

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:00:45 ; Search time 9467.53 Seconds
(without alignments)
10548.289 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061

Sequence: 1 atgcgcagctctctgacga.....ccgtgcgcgcgcgcacatga 2061

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:1:
1: gb ba:*
2: gb hta:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2061	100.0	2061	6 BD178313	BD178313 Transform
2	2061	100.0	2061	6 BD093914	BD093914 Transform
3	2061	100.0	5251	1 AB116234	AB116234 Streptomy
4	1773.8	86.1	2660	1 SVU21728	SVU21728 Streptomy
5	1773.8	86.1	14159	1 AF262220	AF262220 Streptomy
6	820	39.8	2220	6 AR198361	AR198361 Sequence
7	820	39.8	4740	1 SPUG0417	SPUG0417 Streptomy
8	773.4	37.5	4607	1 STMPABA	STMPABA Streptomy
9	773.4	37.5	39314	1 SGR300302	SGR300302 Streptomy
10	769.8	37.4	138203	1 AY310323	AY310323 Streptomy
11	701.6	34.0	29132	1 STH575648	STH575648 Streptomy
12	662.8	32.2	110000	1 AP006618_17	AP006618_17 Streptomy
13	606.2	29.4	298300	1 AP005025	AP005025 Streptomy
14	478.4	23.2	300750	1 AP005217	AP005217 Coryneb
15	459	22.3	4496	6 A48326	A48326 Sequence 4
16	459	22.3	4496	6 AR198356	AR198356 Sequence
17	365.4	17.7	2888	6 A48323	A48323 Sequence 1
18	365.4	17.7	2888	6 AR198353	AR198353 Sequence
19	361.8	17.6	55839	1 AJ586576	AJ586576 Xanthomon

C 20	335.6	16.3	321250	1 SC0939111	AL939111 Streptomy
C 21	323	15.7	276800	1 SC0939115	AL939115 Streptomy
22	317.6	15.4	110000	1 AE017283_21	Continuation (22 o
23	316	15.3	26309	6 CQ363747	CQ363747 Sequence
24	307.4	14.9	110000	1 AE016822_12	Continuation (13 o
25	301.4	14.6	299925	1 AP005045	AP005045 Streptomy
26	294.6	14.3	4367	1 HLFRP	M83788 Haloterrax v
27	282.4	13.7	110000	1 AE017283_30	Continuation (31 o
28	280.4	13.6	340900	1 SM591791	AL591791 Sinorhizo
29	278	13.5	110000	1 AP006840_15	Continuation (16 o
30	272.2	13.2	645	6 A37839	A37839 Sequence 9
31	272.2	13.2	645	6 AR069877	AR069877 Sequence
32	272.2	13.2	645	6 AR099273	AR099273 Sequence
33	272.2	13.2	645	6 AR124158	AR124158 Sequence
34	272.2	13.2	645	6 AR442765	AR442765 Sequence
35	270	13.1	15474	1 AE004601	AE004601 Pseudomon
36	268.2	13.0	348077	1 AP003000	AP003000 Mesorhizo
C 37	267.2	12.9	110000	1 AE017180_26	Continuation (27 o
38	266.8	12.9	1860	6 BD163312	BD163312 Novel pol
39	266.8	12.9	1860	6 AX121195	AX121195 Sequence
40	266.8	12.9	1983	6 AX064177	AX064177 Sequence
41	266.8	12.9	1983	6 AX064419	AX064419 Sequence
42	266.8	12.9	2005	6 AX764213	AX764213 Sequence
43	266.8	12.9	333150	1 AP005277	AP005277 Coryneb
44	266.8	12.9	349459	1 BX927151	BX927151 Coryneb
45	266.8	12.9	349980	6 AX127146	AX127146 Sequence

ALIGNMENTS

RESULT 1	BD178313	2061 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD178313	Transformant producing PF1022 substance and process for producing the same and novel biosynthetic gene.			
DEFINITION	BD178313	Transformant producing PF1022 substance and process for producing the same and novel biosynthetic gene.			
ACCESSION	BD178313	BD178313.1 GI:30015578			
VERSION	BD178313	WO 02077244-A/1.			
KEYWORDS	BD178313	Streptomyces venezuelae			
SOURCE	BD178313	Streptomyces venezuelae			
ORGANISM	BD178313	Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.			
REFERENCE	1 (bases 1 to 2061)	Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.			
AUTHORS	Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.	Transformant producing PF1022 substance and process for producing the same and novel biosynthetic gene			
TITLE	Patent: WO 02077244-A 1 03-OCT-2002;				
JOURNAL	MEIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
COMMENT	TATSUKI MORIYA, TAKESHI MURAKAMI				
OS	Streptomyces venezuelae				
PN	WO 02077244-A/1				
PD	03-OCT-2002				
PF	22-MAR-2002	WO 2002JP002782			
PR	22-MAR-2001	JP OIP 082227			
PI	KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
PC	MURAKAMI				
PC	C12N1/5.61, C12N1/15, C12P17/14// (C12N1/15, C12R1.645), (C12P17/14, C12R1.645)				
CC	Transformant producing PF1022 substance and process for CC				
CC	Producing the same				
CC	and novel biosynthetic gene				
CC	Location/Qualifiers				
CC	Key				
CC	Location/Qualifiers				
CC	FT				
FEATURES	1..2061				
source	Location/Qualifiers				
source	1..2061				
source	/organism="Streptomyces venezuelae"				
source	/mol_type="genomic DNA"				
source	/db_xref="taxon:54571"				
ORIGIN	100.0%; Score 2061; DB 6; Length 2061;				
Query Match	100.0%; Score 2061; DB 6; Length 2061;				

Best Local Similarity 100.0%; Pred. No. 2,2e-158;
Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGGACGCGTTCTGATGACAACTAGACCTGCTCCACCAACCTGTCAGTACATC 60
Db 1 ATGGGACGCGTTCTGATGACAACTAGACCTGCTCCACCAACCTGTCAGTACATC 60
QY 61 GGGAGGACCAACCGGACCAACCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
Db 61 GGGAGGACCAACCGGACCAACCCCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 120
QY 121 CCGGTGAGAGACTTTCAGACGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
Db 121 CCGGTGAGAGACTTTCAGACGAGATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 180
QY 181 CGGGAAGTTGGAAATACACCGCGGCGGATGACAGACAGCGGCGGCGGCGGCGGCGG 240
Db 181 CGGGAAGTTGGAAATACACCGCGGCGGATGACAGACAGCGGCGGCGGCGGCGGCGG 240
QY 241 TGCGTCGGACCAACGAGGACATGCGCAGACTCTTGGGCGGAACCGTGGGCTGGCCG 300
Db 241 TGCGTCGGACCAACGAGGACATGCGCAGACTCTTGGGCGGAACCGTGGGCTGGCCG 300
QY 301 CCCATGCAAGGCGGAGTCTTCAGAGTGTGGCAGACCGGCGAGGACGTCTTTCGGGAG 360
Db 301 CCCATGCAAGGCGGAGTCTTCAGAGTGTGGCAGACCGGCGAGGACGTCTTTCGGGAG 360
QY 361 CCCGTCGGGTTTCAACCGCGGTGCGCTTACCACTCCCTGGCGGCAACCGACTTCG 420
Db 361 CCCGTCGGGTTTCAACCGCGGTGCGCTTACCACTCCCTGGCGGCAACCGACTTCG 420
QY 421 CTGGAACCCCTCGCTGTGAGACGAGACGAGGAGTGTGTGTGTGTGTGTGTGTGTGT 480
Db 421 CTGGAACCCCTCGCTGTGAGACGAGACGAGGAGTGTGTGTGTGTGTGTGTGTGTGT 480
QY 481 CCGCTGTGGGCGGTTCAGTTTCAACCGGAGTCCATGCGGACGCACTTTCGGGAGAT 540
Db 481 CCGCTGTGGGCGGTTCAGTTTCAACCGGAGTCCATGCGGAGATCCTTTCGGGAGAT 540
QY 541 ATGGGCAACTTTCGCGGCGCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 600
Db 541 ATGGGCAACTTTCGCGGCGCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 600
QY 601 TCCCGTACGAACTTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 601 TCCCGTACGAACTTCAAGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
QY 661 CGGCGCTGCTTCCCGGCGGAGGAGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
Db 661 CGGCGCTGCTTCCCGGCGGAGGAGACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 720
QY 721 GGGGCGTGGGCGTTCCTTCTTCGAGGAGACGCGGCGGCGGCGGCGGCGGCGGCGG 780
Db 721 GGGGCGTGGGCGTTCCTTCTTCGAGGAGACGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 781 TACGCGTGTGCGGAGCGGCGTGTCTCCGTCGCGGCGTCCGAGCGACACGACCGGAG 840
Db 781 TACGCGTGTGCGGAGCGGCGTGTCTCCGTCGCGGCGTCCGAGCGACACGACCGGAG 840
QY 841 CGGCGCGCTTCTTCAACTTGTGAGAGACGAGTGTGTGTGTGTGTGTGTGTGTGTGT 900
Db 841 CGGCGCGCTTCTTCAACTTGTGAGAGACGAGTGTGTGTGTGTGTGTGTGTGTGTGT 900
QY 901 CCGGAAGTGGCGTTTCGAGTTCAACTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGT 960
Db 901 CCGGAAGTGGCGTTTCGAGTTCAACTGTGAGTGTGTGTGTGTGTGTGTGTGTGTGT 960
QY 961 GCGGAGACCAACCGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
Db 961 GCGGAGACCAACCGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
QY 1021 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
Db 1021 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
```

```
Db 1081 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1080
QY 1081 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
Db 1081 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1140
QY 1141 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
Db 1141 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1200
QY 1201 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
Db 1201 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1260
QY 1261 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
Db 1261 GCGGACCGGCGCAATCGGCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1320
QY 1321 AACATGATGACGCGCGCGACCGAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 1380
Db 1321 AACATGATGACGCGCGCGACCGAGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCG 1380
QY 1381 ATCAGCCCCGTCGCTGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
Db 1381 ATCAGCCCCGTCGCTGTAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
QY 1441 TGCGCCGAGCGGTTCTTCAAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
Db 1441 TGCGCCGAGCGGTTCTTCAAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1500
QY 1501 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
Db 1501 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
QY 1561 GGGCGGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
Db 1561 GGGCGGAGAGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
QY 1621 AACAGGATGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
Db 1621 AACAGGATGTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1680
QY 1681 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1740
Db 1681 GGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1740
QY 1741 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
Db 1741 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
QY 1801 ACCATGGAATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
Db 1801 ACCATGGAATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
QY 1861 GGAATGATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
Db 1861 GGAATGATGATGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
QY 1921 GCGGACGCGGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
Db 1921 GCGGACGCGGCGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980
QY 1981 GAGGAGTTTCAACGAGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
Db 1981 GAGGAGTTTCAACGAGACCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2040
QY 2041 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2061
Db 2041 GCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2061
```

RESULT 2
BD093914

LOCUS BD093914 2061 bp DNA linear PAT 27-AUG-2002
DEFINITION Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.
ACCESSION BD093914
VERSION BD093914.1 GI:22639502
KEYWORDS MO 0123542-A/1.
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 2061)
Yana, K., Okakura, K., Yasuda, S., Watanabe, M., Miyamoto, K., Mido, N., and Murakami, T.
TITLE Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes
JOURNAL Patent: WO 0123542-A 1 05-APR-2001;
MEIJI SEIKA KAISHA LTD, KOJI YANAI, KOKU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI
COMMENT OS Streptomyces venezuelae
PN MO 0123542-A/1
PD 05-APR-2001
PR 29-SEP-2000 WO 2000JP06783
PR 29-SEP-1999 JP 99 276314
PI KOJI YANAI, KOKU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI
PC C12N15/09, C12N5/10, C12P21/02, C07K11/00//C12P21/02, C12R1.645)
CC
FH
FT CDS
KEY Location/Qualifiers
1. 2061
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
ORIGIN
Query Match 100.0%; Score 2061; DB 6; Length 2061;
Best Local Similarity 100.0%; Pred. No. 2,2e-158;
Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGCGACGCTTGTGATGACAACTAGACCTGTTACCCCAACCTGTCAGTATCATC 60
DB 1 ATGCGACGCTTGTGATGACAACTAGACCTGTTACCCCAACCTGTCAGTATCATC 60
QY 61 GCGAGGCGCACCGGCGAACCCCGGTGTCGTCGCAACGCGGCACTGTGCGGCTG 120
DB 61 GCGAGGCGCACCGGCGAACCCCGGTGTCGTCGTCGCAACGCGGCACTGTGCGGCTG 120
QY 121 CCGGTGAGAGACTTTCGACGCGGATGTCGTGTCGCGGCGCGGCGGCGGCGGCGG 180
DB 121 CCGGTGAGAGACTTTCGACGCGGATGTCGTGTCGCGGCGCGGCGGCGGCGGCGG 180
QY 181 CCGGACTTCGGAATCAAGCCGCGGCGGATGATGACGACGCGGCTGCGGCTGCGGCTC 240
DB 181 CCGGACTTCGGAATCAAGCCGCGGCGGATGATGACGACGCGGCTGCGGCTGCGGCTC 240
QY 241 TGCTCTGGCCACGAGGCGATGCGCCAGCTCTTGGCGGAAACCGTGGCTGCGCGCGAA 300
DB 241 TGCTCTGGCCACGAGGCGATGCGCCAGCTCTTGGCGGAAACCGTGGCTGCGCGCGAA 300
QY 301 CCGATGACGCGCGGCGGCTTCGAGGTGCGGACACCGGCGAGAGCGTTCGCGGCGCTC 360
DB 301 CCGATGACGCGCGGCGGCTTCGAGGTGCGGACACCGGCGAGAGCGTTCGCGGCGCTC 360
QY 361 CCGCTGCGGCTTCAACGCGCGTGTACCACTCCCTGGCGCGACCGACTCTCCCGAGAG 420
DB 361 CCGCTGCGGCTTCAACGCGCGTGTACCACTCCCTGGCGCGACCGACTCTCCCGAGAG 420
QY 421 CTGGAACCCCTGCGCTGAGAGGAGACGCGGCTGTCATGAGGCTGCGGCGACCGGAGAG 480
DB 421 CTGGAACCCCTGCGCTGAGAGGAGACGCGGCTGTCATGAGGCTGCGGCGACCGGAGAG 480

QY 481 CCGCTGAGGCGCTCCAGTTCACCCGAGTCCATCGGACGACTTCGCGCGGAGATC 540
DB 481 CCGCTGAGGCGCTCCAGTTCACCCGAGTCCATCGGACGACTTCGCGCGGAGATC 540
QY 541 ATGGCCAACTTTCGCGACCTCGGCTCGCCTGACACCGGCGACCGGCGGCGGCGAC 600
DB 541 ATGGCCAACTTTCGCGACCTCGGCTCGCCTGACACCGGCGACCGGCGGCGGCGAC 600
QY 601 TCCCGGTAAGAACTTCAAGTGGCGCGGCTGAGAGTGTGCGGACCGCGGAGAGTACG 660
DB 601 TCCCGGTAAGAACTTCAAGTGGCGCGGCTGAGAGTGTGCGGACCGCGGAGAGTACG 660
QY 661 CCGGCGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
DB 661 CCGGCGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 720
QY 721 GCGGCTGCGGCTTTCCTTCTTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
DB 721 GCGGCTGCGGCTTTCCTTCTTCTTCTGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 781 TACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB 781 TACCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840
QY 841 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
DB 841 CCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
QY 901 CCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
DB 901 CCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
QY 961 GCGGAGACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
DB 961 GCGGAGACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020
QY 1021 GCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
DB 1021 GCGGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1080
QY 1081 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
DB 1081 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
QY 1141 GCGCTGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
DB 1141 GCGCTGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1200
QY 1201 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
DB 1201 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1260
QY 1261 CCGATGACGAGTGTCTCAAGAGATCGGCAACCGGAGTGTGATGAGTCTGCGGAC 1320
DB 1261 CCGATGACGAGTGTCTCAAGAGATCGGCAACCGGAGTGTGATGAGTCTGCGGAC 1320
QY 1321 AACATGATCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1380
DB 1321 AACATGATCAACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1380
QY 1381 ATGAGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
DB 1381 ATGAGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1440
QY 1441 TCGCGGAGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
DB 1441 TCGCGGAGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500
QY 1501 GCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
DB 1501 GCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1560
QY 1561 GCGCGGAGAGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620

Db	1561	GGCCGGAGAGAGCGCGGCCGAGAACCTGATGATCTGACCGTGTCCGCAACGACCTC	1620
Qy	1621	AACAGCGTCTGCGCGATCGGCTCCGTCACGTGCCCCGCGCTTTCGAGGTGAGACCTAC	1680
Db	1621	AACAGCGTCTGCGCGATCGGCTCCGTCACGTGCCCCGCGCTTTCGAGGTGAGACCTAC	1680
Qy	1681	GCGCCCGTGCACGAGCTGTGTGTCACCATCCGGGAGCGGTGCGGCCCGGACACGACACC	1740
Db	1681	GCGCCCGTGCACGAGCTGTGTGTCACCATCCGGGAGCGGTGCGGCCCGGACACGACACC	1740
Qy	1741	GCGCCCGTGCACGAGCTGTGTGTCACCATCCGGGAGCGGTGCGGCCCGGACACGACACC	1800
Db	1741	GCGCCCGTGCACGAGCTGTGTGTCACCATCCGGGAGCGGTGCGGCCCGGACACGACACC	1800
Qy	1801	ACCATGAGATGATTCAGACCGCTGAGAGAGGCCCCGGGGCGGTCTACTCCGGGGCGCTC	1860
Db	1801	ACCATGAGATGATTCAGACCGCTGAGAGAGGCCCCGGGGCGGTCTACTCCGGGGCGCTC	1860
Qy	1861	GGATGTGTGCGCCCTCAGCGGGCGCGCGACCTCAGCATCTGATCCGACCATCTGCTG	1920
Db	1861	GGATGTGTGCGCCCTCAGCGGGCGCGCGACCTCAGCATCTGATCCGACCATCTGCTG	1920
Qy	1921	GCCGACGCGCAGGCGGAGTTCCGCGTGGCGGGCGCATCTGTCCTCTCCGACGAGAG	1980
Db	1921	GCCGACGCGCAGGCGGAGTTCCGCGTGGCGGGCGCATCTGTCCTCTCCGACGAGAG	1980
Qy	1981	GAGGAGTTTCACCGAGACCGTGTAAAGCCCGCGCCATGATCACCCTCCGACGAGAG	2040
Db	1981	GAGGAGTTTCACCGAGACCGTGTAAAGCCCGCGCCATGATCACCCTCCGACGAGAG	2040
Qy	2041	GCCGTGGCGGGCGCGCGATGA 2061	
Db	2041	GCCGTGGCGGGCGCGCGATGA 2061	
RESULT 3			
AB116234			
LOCUS	AB116234	5251 bp DNA linear BCT 03-JUL-2004	
DEFINITION	AB116234	Streptomyces venezuelae papA, papC, papB, ORFV genes for 4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyprephenate dehydrogenase, 4-amino-4-deoxychorismate mutase, hypotnetical protein, complete cds.	
ACCESSION	AB116234	GI:47846865	
VERSION	AB116234		
KEYWORDS		Streptomyces venezuelae	
SOURCE		Streptomyces venezuelae	
ORGANISM		Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.	
REFERENCE			
AUTHORS		Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and Murakami, T.	
TITLE		Para-position derivatives of fungal antihelminthic cyclopeptide biosynthetic genes	
JOURNAL		Nat. Biotechnol. 22 (7), 848-855 (2004)	
PUBMED		15184904	
REFERENCE			
AUTHORS		Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and Murakami, T.	
TITLE		Direct Submision	
JOURNAL		Submitted (01-AUG-2003) Koji Yanai, Meiji Seika Kaisha, Ltd., Microbiological Resources and Technology Laboratories; 788 Kayama, Odawara-shi, Kanagawa, 2500852, Japan	
		(E-mail: koji_yanai@meiji.co.jp, Tel: 81-465-37-5106, Fax: 81-465-37-6397)	
FEATURES			
source		Location/Qualifiers	
		1. .5251	
		/organism="Streptomyces venezuelae"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:54571"	
		651. .2711	
gene			

CDS		/gene="papA"	
		651. .2711	
		/gene="papA"	
		/codon_start=1	
		/evidence=experimental	
		/transl_table=11	
		/product="4-amino-4-deoxychorismate synthase"	
		/protein_id="BAD21140.1"	
		/db_xref="GI:47846866"	
		/translation="MRTLLINDVDSPTNHLFOYIGEAATGQPPVVPVNPDAWSRLPYED FPAIVSPGSPDRBDPDISRAITTSGLSVLVGICLHQGIATQFGSTVGLAPPM HGRVSEVRRTGSDVFRGLSPPTAVRYSLSLATIDPLEPLAMSDGVMGLRRREK PLMGVQFHPSIGSDFRGELIMANFRDLALAHRAHRHAGDSPEYLHVRVADLPABE VRRGCLPGEGTTFMLDSSSVLEGCASRFSFLGDRGPLAEYLTTRYADGVSVRSGDT TTRRRPFENYLEBOLERRRVPVAPELPEFNLGVYGLYELKATTTGDDPAHRSPP DAAPLPAADATLDHOCCCYLALDRGHDDGAAWMEPETAETLTGLAVRAPARPT AMVFGPEAAAFGLARRRHDKDAYLKRIDELKEIRNGESYETCLTMVTAPTBAT ALPLYSALPAISPVPGALPELPELSVLSASERELTIGADGVESKPKTIRPRGCT AEDEBRLRADLAGREKRAENLMTIDLVANDLNSVCAIGSVHVPRLFEVETVAPYHOL VSTIRGLRPGTSTAACVAAAPPGGSMGAPKKRTMEIIDRLEEGPRGVSGALCMFA LSGAADLSIVIRTTIVLAADQAEFGVGAIVLSLDQEEFTETVVARAMVMTADGSAV AGAR"	
gene	2708. .3676	/gene="papC"	
CDS	2708. .3676	/gene="papC"	
		/codon_start=1	
		/evidence=experimental	
		/transl_table=11	
		/product="4-amino-4-deoxyprephenate dehydrogenase"	
		/protein_id="BAD21141.1"	
		/db_xref="GI:47846867"	
		/translation="MSGPFRSVVVGSGAGVGMFAGLIREAGRTLVVDVLPVRPD ACTVGDVTPARGPBELAAALRDADVLLAAHEVDALAXVAPYRPMKPGALLADTSVRT GMAAELAAHAPGVQVGLNPNPAPAGMTGRVPAVVTIRDGGVTLALRLVVGSGRP VRLTAEBHDRTTAATQALTHAVLLFGLALALAGVDVRAALATAPRHOVLALLARV LGSPPEVYDIDIRSNPRAASARALAEALRSFPAALVGDPPADAPGRADAGHGGCG DGAGNLDGVFGLRMLMGPBELAAGQDHCOELPRTLRTDDEGEKOR"	
gene	3673. .3984	/gene="papB"	
CDS	3673. .3984	/gene="papB"	
		/codon_start=1	
		/evidence=experimental	
		/transl_table=11	
		/product="4-amino-4-deoxychorismate mutase"	
		/protein_id="BAD21142.1"	
		/db_xref="GI:47846868"	
		/translation="MTEQNELORLRAELDALDGTLLDTVRRRIDLGRIARYKSRHGV PMQPGRAVSLVXDRAARYADHGDSEFLVNLVYIITEMCKVEDLVMSRESLTAEDR R"	
gene	4065. .5243	/gene="ORFIV"	
CDS	4065. .5243	/gene="ORFIV"	
		/codon_start=1	
		/evidence=not experimental	
		/transl_table=11	
		/product="hypothetical protein"	
		/protein_id="BAD21143.1"	
		/db_xref="GI:47846869"	
		/translation="MAATTTLALATGRTAAQOPYPDPDAVDATVTERLAGLPLVLPED EVDIDMGMALVARGAALLVGGDCAEFHEVPLVARRKVDNIOGLAIRASGGLP AVALRIRGQYKRRSSPFETBESAGKMSYCGDAYNEPEFEAATTPERRLLITA YTCGIVIDEIRKRSRGPVLERVYTSHELLLPYERLVNEGAGTSGSAHFHGIC BTRTREDGAHYLAQVANHVPVKLGPVSPEDAVALSRLNEPEVGRLLTFYVFGA KEVDLPLPVVAVARGAPVWMLCDPMHGGLKLAGKTRLIEPMRAETAFAVFTLR EHGQWPAGLHELTPDPTVTCVSELDRRPRFTDYNSCTDPRLNPEQSDMTWTHFLSL "	
ORIGIN			
Query Match	100.0%	Score 2061; DB 1; Length 5251;	
Best Local Similarity	100.0%	Pred. No. 1.5e-158;	

Matches 2061: Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
QY	1	ATGCGACGCTTGTGATCGACAATACTGCTGTTACCCACAACCTGTTGAGTATCATC	60	
Db	651	ATGCGACGCTTGTGATCGACAATACTGCTGTTACCCACAACCTGTTGAGTATCATC	710	
QY	61	GCGGAGGCCACCGGGGACCCCGTGTGTGTCGACGACGACGCTGTGTGCGGGCTG	120	
Db	711	GCGGAGGCCACCGGGGACCCCGTGTGTGTCGACGACGACGCTGTGTGCGGGCTG	770	
QY	121	CCGCTGAGAGACTTTCAGACGAGATGCTGTGTCCCGGGCCCGGACGCCGACCGGAA	180	
Db	771	CCGCTGAGAGACTTTCAGACGAGATGCTGTGTCCCGGGCCCGGACGCCGACCGGAA	830	
QY	181	CGGAGCTTGGGAATCAGCCGCGGGGAGATGACGACGAGCGGCTGTGCGGCTGTGCGGCTC	240	
Db	831	CGGAGCTTGGGAATCAGCCGCGGGGAGATGACGACGAGCGGCTGTGCGGCTGTGCGGCTC	890	
QY	241	TGCGTGGGCAACGAGGAGCATCGCCAGCTTTCGGGCGGAAACGTCGGGCTTCGCCGGA	300	
Db	891	TGCGTGGGCAACGAGGAGCATCGCCAGCTTTCGGGCGGAAACGTCGGGCTTCGCCGGA	950	
QY	301	CCCATGACAGCGCGGGTCTTCGAGGTGCGGACACCGGCGAGGACGTCTTCGCGGGCTTC	360	
Db	951	CCCATGACAGCGCGGGTCTTCGAGGTGCGGACACCGGCGAGGACGTCTTCGCGGGCTTC	1010	
QY	361	CCCTGCGCTTCAACGCGCGGTGCTACCTCCCTGGCGGCAACGACCTTCGCCGACGAG	420	
Db	1011	CCCTGCGCTTCAACGCGCGGTGCTACCTCCCTGGCGGCAACGACCTTCGCCGACGAG	1070	
QY	421	CTCGAACCCCTCGCTGAGACGACGAGGGTGTATGAGGCTTCGGGCAACCGGAGAA	480	
Db	1071	CTCGAACCCCTCGCTGAGACGACGAGGGTGTATGAGGCTTCGGGCAACCGGAGAA	1130	
QY	481	CGGCTGTGGGGCTTCAGATTCCACCCGAGTTCATCGGACGACCTTCGGCGGGAGATC	540	
Db	1131	CGGCTGTGGGGCTTCAGATTCCACCCGAGTTCATCGGACGACCTTCGGCGGGAGATC	1190	
QY	541	ATGGCAACTTCGCGGACCTCGCGCTCGCCACACCGGGGACGAGGGCCGAGCGGAC	600	
Db	1191	ATGGCAACTTCGCGGACCTCGCGCTCGCCACACCGGGGACGAGGGCCGAGCGGAC	1250	
QY	601	TCCCGCTAGAACTCCACGTCGCGCGCTGACGTCGTGCGGACCGGAAAGATACCG	660	
Db	1251	TCCCGCTAGAACTCCACGTCGCGCGCTGACGTCGTGCGGACCGGAAAGATACCG	1310	
QY	661	CGGCGCTGTGCTCGCGGACGACCAAGTTCGTGCTGACGACGCTTCGCTCGAA	720	
Db	1311	CGGCGCTGTGCTCGCGGACGACCAAGTTCGTGCTGACGACGCTTCGCTCGAA	1370	
QY	721	GCGGCTGTGCTTCCTTCCTTCGCGGACGACCGCGGCGCTTCGCGGATACCTCAC	780	
Db	1371	GCGGCTGTGCTTCCTTCCTTCGCGGACGACCGCGGCGCTTCGCGGATACCTCAC	1430	
QY	781	TACCGGTCGCGGACGAGCGCTGTCTCGTCCGCGGCTCGGACGCGGACGACGAGCGGAG	840	
Db	1431	TACCGGTCGCGGACGAGCGCTGTCTCGTCCGCGGCTCGGACGCGGACGACGAGCGGAG	1490	
QY	841	CGGCGCGCTTCTTCACTACCTGAGGAGACAGCTCGAAACGCGGACGCGGCTCGCG	900	
Db	1491	CGGCGCGCTTCTTCACTACCTGAGGAGACAGCTCGAAACGCGGACGCGGCTCGCG	1550	
QY	901	CCCGAACTGCGCTTTCAGATTCAACCTCGGCTACGTGAGGCTACCTCGGCTACGAGCTGA	960	
Db	1551	CCCGAACTGCGCTTTCAGATTCAACCTCGGCTACGTGAGGCTACCTCGGCTACGAGCTGA	1610	
QY	961	GCGGAGACGAGCGGACCGCGGACGAGTTCGCGGACCGCGGACGCGGCTTCCTTC	1020	
Db	1611	GCGGAGACGAGCGGACCGCGGACGAGTTCGCGGACCGCGGACCGCGGCTTCCTTC	1670	
QY	1021	GCGGACGCGGCACTGCGCTTCGACCAACGAGAGGCTGTGCTACCTGTGCGCTTCGAC	1080	
Db	1671	GCGGACGCGGCACTGCGCTTCGACCAACGAGAGGCTGTGCTACCTGTGCGCTTCGAC	1730	

QY	1081	CGCGGGGGCCACGACGAGCGGCGCGGCTGTGCGGAGACGAGCGGACCCCTCAC	1140	
Db	1731	CGCGGGGGCCACGACGAGCGGCGCGGCTGTGCGGAGACGAGCGGACCCCTCAC	1790	
QY	1141	GCGCTGTGCGGCGCGCGCGCGGACCGGACCGCGGCTGTGCGGATTCGCGAG	1200	
Db	1791	GCGCTGTGCGGCGCGCGCGCGGACCGGACCGCGGCTGTGCGGATTCGCGAG	1850	
QY	1201	GCGGCGGCGGCTTGTGCGCGCGCGCGGACCGGACCGGACCGGCTGTGCGGATTCGCGAG	1260	
Db	1851	GCGGCGGCGGCTTGTGCGCGCGCGGACCGGACCGGACCGGCTGTGCGGATTCGCGAG	1910	
QY	1261	CGCATGACGAGTGTCTCAAGAGATCGGCAACGCGGAGTGTGAGAGTTCGCGGAC	1320	
Db	1911	CGCATGACGAGTGTCTCAAGAGATCGGCAACGCGGAGTGTGAGAGTTCGCGGAC	1970	
QY	1321	AACATGTGACCGGCGGACCGGAGGAGCGGCGGCTGTGCTGTACTCCGCGCTGCGGCG	1380	
Db	1971	AACATGTGACCGGCGGACCGGAGGAGCGGCGGCTGTGCTGTACTCCGCGCTGCGGCG	2030	
QY	1381	ATCAGCCCGTCCGTAACGCGCGCTGTGAGTTCCCGAACTGTGCTGAGCGCC	1440	
Db	2031	ATCAGCCCGTCCGTAACGCGCGCTGTGAGTTCCCGAACTGTGCTGAGCGCC	2090	
QY	1441	TGCGCGGAGCGGCTTCTCAAGATCGGCGCGGACCGGCGGCTGTGAGTTCGAGCGCAATCA	1500	
Db	2091	TGCGCGGAGCGGCTTCTCAAGATCGGCGCGGACCGGCGGCTGTGAGTTCGAGCGCAATCA	2150	
QY	1501	GCGACCGCGCGCGGCGGACCGGAGGAGGACGAGCGGCTCGGCGGACCGGCTGCGC	1560	
Db	2151	GCGACCGCGCGCGGCGGACCGGAGGAGGACGAGCGGCTCGGCGGACCGGCTGCGC	2210	
QY	1561	GCGCGGAGAGGACCGGCGGAGGACCTGAATGTGTGACCTGTGCTGCGCAACGACTTC	1620	
Db	2211	GCGCGGAGAGGACCGGCGGAGGACCTGAATGTGTGACCTGTGCTGCGCAACGACTTC	2270	
QY	1621	AAACAGTGTGCGGATCGGCTTCGCTGACGCGGCGGCTGTGAGTGTGAGGACCTTC	1680	
Db	2271	AAACAGTGTGCGGATCGGCTTCGCTGACGCGGCGGCTGTGAGTGTGAGGACCTTC	2330	
QY	1681	GCGCGGTCACCAAGTGTGTGACCAATCGGAGGACGAGCTGCGGCGGACCGGAC	1740	
Db	2331	GCGCGGTCACCAAGTGTGTGACCAATCGGAGGACGAGCTGCGGCGGACCGGAC	2390	
QY	1741	GCGCGCTGTGACGCGCGGCTTTCGCGGCGGCTTCATGACGCGGCGGCAAGGCGC	1800	
Db	2391	GCGCGCTGTGACGCGCGGCTTTCGCGGCGGCTTCATGACGCGGCGGCAAGGCGC	2450	
QY	1801	ACCATGAGATATGACGCGCTGTGAGGAGGACCGGCGGCTGTGCTGCGGCGGCTTC	1860	
Db	2451	ACCATGAGATATGACGCGCTGTGAGGAGGACCGGCGGCTGTGCTGCGGCGGCTTC	2510	
QY	1861	GATGATGCTGCGCTGAGCGGCGGCGGACCTCAGATGTGATCGGACCAATCGGCTG	1920	
Db	2511	GATGATGCTGCGCTGAGCGGCGGCGGACCTCAGATGTGATCGGACCAATCGGCTG	2570	
QY	1921	GCGGACGCGGACGAGTTCGCGCTGTGCGGAGGACCGGCGGCTGTGCTGCGGACGAG	1980	
Db	2571	GCGGACGCGGACGAGTTCGCGCTGTGCGGAGGACCGGCGGCTGTGCTGCGGACGAG	2630	
QY	1981	GAGGAGTTCACGAGACCGTGTGTAAGGCGGCGGCTGTGATCGGCTGTGAGCGGAC	2040	
Db	2631	GAGGAGTTCACGAGACCGTGTGTAAGGCGGCGGCTGTGATCGGCTGTGAGCGGAC	2690	
QY	2041	GCGGTGCGGCGGCGGATGA 2061		
Db	2691	GCGGTGCGGCGGCGGATGA 2711		

RESULT 4
SVU21728
LOCUS

SVU21728

2660 bp

DNA

linear

BCT 22-NOV-1996

DEFINITION Streptomyces venezuelae p-aminobenzoic acid synthase (pabAB) gene,
complete cds.
ACCESSION U21728
VERSION U21728.1 GI:1680048
KEYWORDS
SOURCE
ORGANISM Streptomyces venezuelae
Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE
AUTHORS 1 (bases 1 to 2660)
TITLE Brown, M.P., Aidoo, K.A. and Vining, L.C.
A role for pabAB, a p-aminobenzoate synthase gene of Streptomyces
venezuelae ISP5230, in chloramphenicol biosynthesis
Microbiology 142 (Pt 6), 1345-1355 (1996)
JOURNAL
MEDLINE 96262706
PUBMED 8704974
REFERENCE 2 (bases 1 to 2660)
AUTHORS Brown, M.P.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1995) Biology, Dalhousie University, Halifax,
Nova Scotia B3H 4H6, Canada
COMMENT On Nov 21, 1996 this sequence version replaced gi:1680040.
FEATURES
source
1..2660
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ISP5230"
/db_xref="taxon:54571"
<1..507
/note="ORF1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAB30311.1"
/db_xref="GI:1680049"
/translation="GSAEYVALPFLGEHDLIRSKTGLIRNGERSVLPADSTNIS
PMTYKVAEIVGPDVTFIGMESIGAAAMTIGPLGEPIDRDTDSRLNSNFOQA
REIVDLPEDEVYVYAMGLEPMWGVMAVDYDESHPAIVDSLIVRHVDKGTERTL
HURRLRL"
651..2660
/gene="pabAB"
651..2660
/gene="pabAB"
/codon_start=1
/transl_table=11
/product="p-aminobenzoic acid synthase"
/protein_id="AAB30312.1"
/db_xref="GI:710438"
/translation="MRTLLIDNYSFTONI,FOYIGRATGQPPVNPADWSRLPLEDF
DAIVSPGSGSPDRERFGISRAITDSGLPVLGVCIGHQIAQLASAPMGRVSEVR
HGEDVPRGLSPFTAVRYSLSATDLPDELPIASDDGVVGLASHREKPLMGVQFP
PESIGDFGREITMANFRDLAHHRRAPADMDGYELHVRVDVLPDAEVRRAACPA
EGATFWLSSSVLEGASPSPLGDRGRLAAYITTVAVANGVSVRSDDTTRDAATL
FYLEQLPBPAPVAPDLPEFNLGTVGLGELKAETTGDPVAPVAPDPDAFLPAD
RAIALDHQGCCYLLALDRGHDDGARAWLRETAETLTGLAVRVRPVPVAVGVPE
AAGFGPLARHDKDASALRNESYEICITNVTAPTBATLPLSALRRLSPVSG
ALLEPELSELVLSPERFLTGADGVSEKPIKGTAPRGAPABEDERLRADLAGREKD
RAENLMTIVLVRNDSVCAIGSVHVPRLFEVGDLPVHOLVSTTGRRLRPPTSTRAK
VAAAPGSGMTGAPKRPMEIITDRLREGRCVLPGLMGPALSGAADLSIVRTIYLA
DGRAERFGVGAIVLSIDQEEERQIVVKARAVTALDGSAAVGAR"

Query Match 86.1%; Score 1773.8; DB 1; Length 2660;
Best Local Similarity 92.8%; Pred. No. 3.2e-135;
Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;

QY 1 ATGCGAGCGCTTGATGACAACTACGACTGTTGACCCACAACCTGTTCCAGTACATC 60
DB 651 ATGCGAGCGCTTGATGACAACTACGACTGTTGACCCACAACCTGTTCCAGTACATC 710
QY 61 GGGGAGGCAACCGGAGCAACCCCGCTGCTGCTGCCCAAGAGCGGACGCTGCTGGCGCTG 120
DB 711 GGGGAGGCAACCGGAGCAACCCCGCTGCTGCTGCCCAAGAGCGGACGCTGCTGGCGCTG 767

QY 121 CCGGTGAGAGACTTGCAGACCGGATCGTGTGTCGCCGGGCCCCGGGACGCCGAGAA 180
DB 768 CCGGTGAGAGACTTGCAGACCGGATCGTGTGTCGCCGGGCCCCGGGACGCCGAGAA 827
QY 181 CGGGAATTGGAATTCAGCCGCCGGGCGATACCGACAGCGGCGCTGCTCGGCGTCT 240
DB 828 CGGGAATTGGAATTCAGCCGCCGGGCGATACCGACAGCGGCGCTGCTCGGCGTCT 887
QY 241 TGCCTGGGACACGAGGGAATCCCGACGCTTTCGGGGAAACCGTGGGCTCCCGCGAA 300
DB 888 TGCCTGGGACACGAGGGAATCCCGACGCTT-----CGGGGAA 926
QY 301 CCCATCAGGCGCGGATCTCCGAGTGGCGGACACCGCGAGGAGACGTTCTTCCGGGAGCTC 360
DB 927 CCCATCAGGCGCGGATCTCCGAGTGGCGGACACCGCGAGGAGACGTTCTTCCGGGAGCTC 986
QY 361 CCCTGCGGCTTCAACCGCGCTGCGCTACCTCCGCGGACCGACCTCCCGACGAG 420
DB 987 CCCTGCGGCTTCAACCGCGCTGCGCTACCTCCGCGGACCGACCTCCCGACGAG 1046
QY 421 CTGGAACCCCTGCGCTGAGCGGACGAGCGGAGTCTCATGAGGCTTGGGCTGCGGACGAG 480
DB 1047 CTGGAACCCCTGCGCTGAGCGGACGAGCGGAGTCTCATGAGGCTTGGGCTGCGGACGAG 1106
QY 481 CCGCTGAGGAGGCTCCAGTTCCACCGGAGTCCATCGGACGAGCACTTGGCGGAGATC 540
DB 1107 CCGCTGAGGAGGCTCCAGTTCCACCGGAGTCCATCGGACGAGCACTTGGCGGAGATC 1166
QY 541 ATGCGCAATTCCGCGACTGCGCTTCCGCGGACGAGCGGCGACCGGCGCGCGGCGGAC 600
DB 1167 ATGCGCAATTCCGCGACTGCGCTTCCGCGGACGAGCGGCGCGCGGCGGAC 1226
QY 601 TCCCGTAGAATCTCACGCTGGCGCGCTGCGAGCTGCGCGGACCGGACGAGAGTACGCC 660
DB 1227 TGGGCTAGAACTCCACGCTGGCGCGCTGCGAGCTGCGCGGACCGGACGAGAGTACGCC 1286
QY 661 CCGGCTGCTGCTCCCGGCGAGGAGCAACAAGTCTGAGTGAAGAGAGCTCCGCTCGAA 720
DB 1287 CCGGCTGCTGCTCCCGGCGAGGAGCAACAAGTCTGAGTGAAGAGAGCTCCGCTCGAA 1346
QY 721 GCGGCTGCTGCTTCTTCTTCTTCTGCGGACGAGCGGCGCGCTGCGCGAGTACTCACC 780
DB 1347 GCGGCTGCTGCTTCTTCTTCTTCTGCGGACGAGCGGCGCGCTGCGCGAGTACTCACC 1406
QY 781 TACCGGTGCGGAGCGGCTGCTTCCGCTGCGGCTCGAGGAGCAACCGACCGCGAGC 840
DB 1407 TACCGGTGCGGAGCGGCTGCTTCCGCTGCGGCTCGAGGAGCAACCGACCGCGAGC 1466
QY 841 CCGGCGCTTCTTCAACTACCTGAGGAGAGAGCTCGAAAGCGCGACCGGATCCCGTGGC 900
DB 1467 CCGGCGCTTCTTCAACTACCTGAGGAGAGAGCTCGAAAGCGCGGAGGATCCCGTGGC 1526
QY 901 CCGGAATGCTTCAAGTTCAACTCGGCTACGTCGAGTACCTCGGCTACGAGCTGAAG 960
DB 1527 CCGGAATGCTTCAAGTTCAACTCGGCTACGTCGAGTACCTCGGCTACGAGCTGAAG 1586
QY 961 GGGGAGCAACCGGAGACCGCGGAGACCGGTCGCGGACCGCGAGCGCGGCTTCTTTC 1020
DB 1587 GGGGAGCAACCGGAGACCGCGGAGACCGGTCGCGGACCGCGAGCGCGGCTTCTTTC 1646
QY 1021 GCGGACCGGCGCATGCGCTTCAACCAACGAGGAGGCTGCTTACCTGTCGCGCTCGAC 1080
DB 1647 GCGGACCGGCGCATGCGCTTCAACCAACGAGGAGGCTGCTTACCTGTCGCGCTCGAC 1706
QY 1081 CCGCGGAGCAAGAGACGAGCGGCGCGCTGCTGCTGCGGAGACGCGGAGACCTCTACC 1140
DB 1707 CCGCGGAGCAAGAGACGAGCGGCGCGCTGCTGCTGCGGAGACGCGGAGACCTCTACC 1766
QY 1141 GCGCTGCGCTTCCGCGCGCGCGGCGAGCGGACCGCGGAGCTTGGGATCCCGGAG 1200
DB 1767 GCGCTGCGCTTCCGCGCGCGCGGCGAGCGGACCGCGGAGCTTGGGATCCCGGAG 1826

QY	1201	GCGGCGGCGGCTTCGGGCCCCCTGCGCGCGCCGCAAGACGACCTTACCTCAAG	1260	REFERENCE	1 (bases 1 to 14159)
Db	1827	GGGGCGGCGGCTTCGGGCCCCCTGCGCGCGCCGCAAGACGACCTTACCTCAAG	1878	AUTHORS	He,J., Magarvey,N., Pirrae,M. and Vining,L.C.
QY	1261	CGCATCGACGAGTCTCTCAAGAGATTCGCAACGCGGAGTGTGAGATTCGCTGACC	1320	TITLE	The gene cluster for chloramphenicol biosynthesis in <i>Streptomyces venezuelae</i> ISP5230 includes novel shikimate pathway homologues and a monomolecular non-ribosomal peptide synthetase gene
Db	1879	-----CGGCGCTCCGCAACGCGGAGTGTGAGATTCGCTGACC	1919	JOURNAL	Microbiology 147 (Pt 10), 2817-2829 (2001)
QY	1321	AACATGGTCAACCGGCGGCGGAGCGACCGGCGGCGGCTTCGCTTACTCGGCGCGCC	1380	REFERENCE	2 (bases 1 to 14159)
Db	1920	AACATGGTCAACCGGCGGCGGAGCGGCGGCGGCTTCGCTTACTCGGCGCGCC	1979	AUTHORS	He,J., Magarvey,N.A. and Vining,L.C.
QY	1381	ATCAGCCCGCTCCCGTACGCGGCGGCTTCGAGTTCGCGGAGTGTGAGGCGCC	1440	TITLE	Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford St., Halifax, NS B3H 4J1, Canada
Db	1980	ATCAGCCCGCTCCCGTACGCGGCGGCTTCGAGTTCGCGGAGTGTGAGGCGCC	2039	JOURNAL	3 (bases 1 to 14159)
QY	1441	TCCGCGGAGCGGCTTCCTACAGATCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	1500	REFERENCE	He,J., Magarvey,N.A. and Vining,L.C.
Db	2040	TCCGCGGAGCGGCTTCCTACAGATCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2099	AUTHORS	Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford St., Halifax, NS B3H 4J1, Canada
QY	1501	GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	1560	TITLE	Sequence update by submitter
Db	2100	GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2159	REMARK	On Jun 5, 2001 this sequence version replaced gi:10716943.
QY	1561	GGCGGAGAGAGAGACCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	1620	COMMENT	location/Qualifiers
Db	2160	GGCGGAGAGAGAGACCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2219	FEATURES	
QY	1621	AACAGCGTCTGCGGAGTCCGCTCCGCTCAGCTGCGGCGGCTTCGAGTCAAGCCCATCAAG	1680	source	1. 14159
Db	2220	AACAGCGTCTGCGGAGTCCGCTCCGCTCAGCTGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2279	CDS	/organism="Streptomyces venezuelae"
QY	1681	GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	1740		/mol_type="genomic DNA"
Db	2280	GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2339		/db_xref="taxon:54571"
QY	1741	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	1800		/map="adjacent to pabAB"
Db	2340	GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGAGTCAAGCCCATCAAG	2399		2333. 3511
QY	1801	ACCATGAGATTCATCGACCGCTTCGAGAGAGAGCGGCGGCGGCTTCGCTCCGCGGCGCTC	1860		/function="essential for chloramphenicol biosynthesis"
Db	2400	ACCATGAGATTCATCGACCGCTTCGAGAGAGAGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2459		/codon_start=1
QY	1861	GATGATTCGCGCTTCGAGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	1920		/cranal_table=11
Db	2460	GATGATTCGCGCTTCGAGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2519		/product="CmlH"
QY	1921	GCGGAGCGGCGGAGTTCGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	1980		/protein_id="AAG21975.2"
Db	2520	GCGGAGCGGCGGAGTTCGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2579		/db_xref="GI:14327897"
QY	1981	GAGGAGTTCAACGAGACCGTGTAAAGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2040		/translation="MELPDLITDLPFRHGRGRTALRTAGTWTYEEIDRVTSALARRID
Db	2580	GAGGAGTTCAACGAGACCGTGTAAAGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2639		ABCPAGRRVLVAGEHTAEVVAALAMRSHAVHTPNPOLPADRFEEFARVADALIV
QY	2041	GCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2061		CEERALVGERAGALALYAGDVGPPTDPAPADGTADAPARSVA51FTSGTGCD
Db	2640	GCGGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTTCGCTCCGCGGCGCTC	2660		PKLVGVGGGLNLCRLRLDLPDDOVLHPASLSPASVSEILGTLYAGATIVVP
RESULT 5					VNDQASWGLSVSRHLLAAGCDLMLSPSYAALDDEAARRIKRKEVCEGALSGEDK
LOCUS	AF262220	14159 bp	DNA	linear	AAIRSRVNAVGPTEATVCSLAEELSYTP5IGTPVDFRAYVRDSDHATGGR
DEFINITION	Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,				LVIIVGVVALGYAGSAPAEVEFGTVDSPAVATGVVLSDDGELTYIGRIDEQIKR
partial sequence.					LGRVNLAVHGSITLSRHLGRVALREIADATILVPAADGAEATESIMARIRDLVPVME
ACCESSION	AF262220				APDRVLVLDALPLTSGKVDRAVALREIADATILVPAADGAEATESIMARIRDLVPVME
VERSION	AF262220.2	GI:14290414			GOEIGPETSIFDAGSSILMIQIOVLTSDVAGEAEVPAALAMDODFAVPLAFLRHGR
KEYWORDS					BAVAPAESAVDTLLRVERTERDALAEELPLARDTRHEVPAADODREVALTLGASGF
SOURCE	Streptomyces venezuelae				IGGVLDRLAAGRPVLVSTGDPGCVLTGHAITREKORADYARVAISVABEUREVD
ORGANISM	Streptomyces venezuelae				RRRGPVDAVHACGYOVNHLPLDLSHLSGSVNTALVVAALAGLPGVYVGHARFEDRLH
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				ADFLPLASATLTVAGDVPYSRKLISEEYVNTLAVGCAVSHRPGILVYGHARFEDRLH
	Streptomyces; Streptomycetaceae; Streptomyces.				KDMPFALAEARVYCAMPLSGHVPVCVGTIADTLGLRPPANPOTADGASRTPEPSR
					SAVVHRTVALDEILRHNTGLTEADVLAAPAEFRVYRDSGEVPAPILAAMQALSGCGM
					PSAHRVDDIIGRLGTRPDPADRPERRCTTSRGPRRRPRADTLTGTRP"
					3576. 3596
					/note="Region: adenylation domain A1"
					3708. 3746
					/note="Region: adenylation domain A2"
					3939. 3986
					/note="Region: adenylation domain A3"

misc_feature 4089..4100
/note="Region: adenylation domain A4"
misc_feature 4350..4370
/note="Region: adenylation domain A5"
misc_feature 4494..4538
/note="Region: adenylation domain A6"
misc_feature 4593..4610
/note="Region: adenylation domain A7"
misc_feature 4647..4706
/note="Region: adenylation domain A8"
misc_feature 4833..4853
/note="Region: adenylation domain A9"
misc_feature 4893..4910
/note="Region: adenylation domain A10"
misc_feature 5043..5075
/note="Region: thiolation domain"
misc_feature 5343..5375
/note="Region: NADH binding site"

ORIGIN

Query Match 86.1%; Score 1773.8; DB 1; Length 14159;
Best Local Similarity 92.8%; Pred. No. 1.6e-135;
Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;

QY 1 ATGCGCAGCTTTCGATCGAACAATACGACTCGTTCAACCAACTCTTTCAGTACATC 60
DB 8153 ATGCGCAGCTTTCGATCGAACAATACGACTCGTTCAACCAACTCTTTCAGTACATC 8212
QY 61 GGGAGGACCAACCGGGCAACCCCGGTGTCGTCGCAAGACGCGACCTGTCGCGGCTG 120
DB 8213 GGGAGGACCAACCGGGCAACCCCGGTGTCGTCGCAAGACGCGACCTGTCGCGGCTG 8269
QY 121 CCCGTGAGGACTTTCGACGCGATGTCGTGTCGCCCGGCCCGGACCCCGACCGGAGAA 180
DB 8270 CCCGTGAGGACTTTCGACGCGATGTCGTGTCGCCCGGCCCGGACCCCGACCGGAGAA 8329
QY 181 CGGAGCTTCGGAATCAAGCCCGCGGCGATACCGACACGCGCTGCCCGTCTTGCGGCTC 240
DB 8330 CGGAGCTTCGGAATCAAGCCCGCGGCGATACCGACACGCGCTGCCCGTCTTGCGGCTC 8389
QY 241 TGCGTCGGGACCAACGAGGACATGCGGACGCTTTCGCGGAGAACGTCGCGCTGCCCGGAGAA 300
DB 8390 TGCGTCGGGACCAACGAGGACATGCGGACGCTTTCGCGGAGAACGTCGCGCTGCCCGGAGAA 8428
QY 301 CCCATGACGCGCGGCTCTCCAGGTGCGACACCGCGGAGACGTCCTTCCGCGGCTC 360
DB 8429 CCCATGACGCGCGGCTCTCCAGGTGCGACACCGCGGAGACGTCCTTCCGCGGCTC 8488
QY 361 CCCCTCGCGTTTCAACCGCGCTGCGCTACACTCTCCGCGCCGACCGACTCTCCCGACGAG 420
DB 8489 CCCCTCGCGTTTCAACCGCGCTGCGCTACACTCTCCGCGCCGACCGACTCTCCCGACGAG 8548
QY 421 CTGGAACCCCTGCGCTGAGAGGACGAGCGGCTGTACATGGGCGCTGGGACCGCGGAGAG 480
DB 8549 CTGGAACCCCTGCGCTGAGAGGACGAGCGGCTGTACATGGGCGCTGGGACCGCGGAGAG 8608
QY 481 CCGCTGAGGCGCTTCAGTTCAACCGGAGTTCATCGGACGCACTTCGCGCGGAGATC 540
DB 8609 CCGCTGAGGCGCTTCAGTTCAACCGGAGTTCATCGGACGCACTTCGCGCGGAGATC 8668
QY 541 ATGGCCAACTTCGCGCACTCGCGCTTCCGACCAACCGGACCGGCGGACCGGCGGAC 600
DB 8669 ATGGCCAACTTCGCGCACTCGCGCTTCCGACCAACCGGACCGGCGGACCGGCGGAC 8728
QY 601 TCCCGTAGGAATCCACGTCGCGCGCGGTGAGAGTGTGTCGCGGACCGGACGAGAGTACGC 660
DB 8729 TCCCGTAGGAATCCACGTCGCGCGCGGTGAGAGTGTGTCGCGGACCGGACGAGAGTACGC 8788
QY 661 CGCGGCTGCTTGCCTCGCGGACGAGCAACGTTTGTGCTGAGACAGAGCTCCGCTCGAA 720
DB 8789 CGCGGCTGCTTGCCTCGCGGACGAGCAACGTTTGTGCTGAGACAGAGCTCCGCTCGAA 8848
QY 721 GCGCGCTCGCGCTTTCCTTCTTCGCGGACGAGCCGCGGCTCGCGGAGTACCTTCACC 780

DB 8849 GCGGCTCGCGCTTTCCTTCTTCGCGGACGAGACCGCGGCTCCGCGGAGTACCTTCACC 8908
QY 781 TACCGGTGCGCGAGCGGCGCTTCCTCCGCGGCTCCGAGGACCAACGACCGGAGG 840
DB 8909 TACCGGTGCGCGAGCGGCGCTTCCTCCGCGGCTCCGAGGACCAACGACCGGAGG 8968
QY 841 CCGCGCGCTTTCCTTCAACTACCTGAGAGAGAGTGAAGAGCGGACCGGCTCCGCTCGC 900
DB 8969 CCGCGCGCTTTCCTTCAACTACCTGAGAGAGAGTGAAGAGCGGACCGGCTCCGCTCGC 9028
QY 901 CCCGAACTGCTTCGAGTTCAACTCTGCGCTACCTGCGCTTACCTCGGCTTACGAGCTGAAG 960
DB 9029 CCCGAACTGCTTCGAGTTCAACTCTGCGCTACCTGCGCTTACCTCGGCTTACGAGCTGAAG 9088
QY 961 GCGGAGACCAACGCGGACCGCGCGGACCGGCTCCGACCGGACCGGCGGCTTCCTTC 1020
DB 9089 GCGGAGACCAACGCGGACCGCGCGGACCGGCTCCGACCGGACCGGCGGCTTCCTTC 9148
QY 1021 GCGGACCGGCGCATGCTTCGACACCAAGAGGCTGCTTACCTGCTGCGCTTCGAC 1080
DB 9149 GCGGACCGGCGCATGCTTCGACACCAAGAGGCTGCTTACCTGCTGCGCTTCGAC 9208
QY 1081 CCGCGGCGCAACGAGCGGCGCGGCTGCTGCTGCGGAGACGCGGACCGTCAAC 1140
DB 9209 CCGCGGCGCAACGAGCGGCGCGGCTGCTGCTGCGGAGACGCGGACCGTCAAC 9268
QY 1141 GCGCTGCGGCTTCGCGGCGCGGCGGACCGGCGGCTTCGCTTTCGAGATTCGCGGAG 1200
DB 9269 GCGCTGCGGCTTCGCGGCGCGGCGGACCGGCGGCTTCGCTTTCGAGATTCGCGGAG 9328
QY 1201 GCGGCGGCGGCTTCGCGGCGCGGCGGACCGGCGGCTTCGCTTTCGAGATTCGCGGAG 1260
DB 9329 GCGGCGGCGGCTTCGCGGCGCGGCGGACCGGCGGCTTCGCTTTCGAGATTCGCGGAG 9380
QY 1261 CGGATGACGAGTGCCTCAAGAGATCCGCAACCGGAGTCTGAGAGATTCGCTGACC 1320
DB 9381 -----CGGCGCTCCGCAACCGGAGTCTGAGAGATTCGCTGACC 9421
QY 1321 AACATGTCACCGCGCGGACCGAGGAGGAGCGGCTGCGCTTACCTCGCGGCTGCGGCGC 1380
DB 9422 AACATGTCACCGCGCGGACCGAGGAGGAGCGGCTGCGCTTACCTCGCGGCTGCGGCGC 9481
QY 1381 ATCAGCCCGCTCCGTAACGCGCGCTGCTGAGTTCCCGGAACTGTGCTGAGCGCGC 1440
DB 9482 ATCAGCCCGCTCCGTAACGCGCGCTGCTGAGTTCCCGGAACTGTGCTGAGCGCGC 9541
QY 1441 TGCGCGAGCGGTTCTTCAAGATCGGCGCGGACGCGGCTGAGTCCAGGCCATCAAG 1500
DB 9542 TGCGCGAGCGGTTCTTCAAGATCGGCGCGGACGCGGCTGAGTCCAGGCCATCAAG 9601
QY 1501 GGGACCGCGCGCGGCGGCGGACCGGAGGAGGAGCGGCTCCGCGCGGACCGGCTGCGC 1560
DB 9602 GGGACCGCGCGCGGCGGCGGACCGGAGGAGGAGCGGCTCCGCGCGGACCGGCTGCGC 9661
QY 1561 GCGCGGAGAGAGACCGGCGCGGACCGTATGATGTCGACTGTGCTGCTGCAACGACTTC 1620
DB 9662 GCGCGGAGAGAGACCGGCGCGGACCGTATGATGTCGACTGTGCTGCTGCAACGACTTC 9721
QY 1621 AACAGGTTTGGCGGATCGGCTTCCTTCCGTCACGTCGCGCGGCTTTCGAGGTGAGACTAC 1680
DB 9722 AACAGGTTTGGCGGATCGGCTTCCTTCCGTCACGTCGCGCGGCTTTCGAGGTGAGACTAC 9781
QY 1681 GCGCGGTCACAGTGCTGTGACCAATCGGAGGAGCGGCTGCGGCGGCGGACCGAGCAC 1740
DB 9782 GCGCGGTCACAGTGCTGTGACCAATCGGAGGAGCGGCTGCGGCGGCGGACCGAGCAC 9841
QY 1741 GCGCGCTGCGTACGCGCGCTTTCGCGGCGGCTTCATGACCGGCGGCGGCGGCGGCGG 1800
DB 9842 GCGCGCTGCGTACGCGCGCTTTCGCGGCGGCTTCATGACCGGCGGCGGCGGCGGCGG 9901
QY 1801 ACCATGAGATATGACCGGCTTGAAGAGGCGCGGCGGCTTACTTCGCGGCGGCTC 1860

Db 9902 CCCATGAGATCATGACCCGCTGAGAGAAAGCCCCCGGGCGTCTTAACCGGGGGCCTC 9661
Qy 1861 GGATGTTTCGCTCAAGGGGGCGCGGACCTGACATCTGTCATCCGCAACATGTCCTG 1920
Db 9962 GGATGTTTCGCTCAAGGGGGCGCGGACCTGACATCTGTCATCCGCAACATGTCCTG 10021
Qy 1921 GCCGACGGGCGGAGGAGTTTCGCGTCCGCGGGGCGATCTGTCCTCTCCGACCAAGAG 1980
Db 10022 GCCGACGGGCGGAGGAGTTTCGCGTCCGCGGGGCGATCTGTCCTCTCCGACCAAGAG 10081
Qy 1981 GAGGATTCACCGAGACCTGTGTAAAGGCCCGCCCATGTGTACCGGCTTCGACGGAGC 2040
Db 10082 GAGGATTCACCGAGACCTGTGTAAAGGCCCGCCCATGTGTACCGGCTTCGACGGAGC 10141
Qy 2041 GCCGTGGCGGGCGCCCGCATGA 2061
Db 10142 GCAGTGGCGGGCGCCCGCATGA 10162

RESULT 6

AR198361 2220 bp DNA linear PAT 20-APR-2002
LOCUS AR198361
DEFINITION Sequence 14 from patent US 6352839.
ACCESSION AR198361
VERSION AR198361.1 GI:20248210
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2220)
AUTHORS Blanc,V., Thibaut,D., Bames-Jacques,N., Blanche,F., Crouzet,J.,
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and
Dutric-Robert,G.
Strepogramins for preparing same by mutasynthesis
Patent: US 6352839-A 14 05-MAR-2002;
Location/Qualifiers
1..2220
/organism="unknown"
/mol_type="unassigned DNA"

TITLE
JOURNAL
FEATURES
source

ORIGIN

Query Match 39.8%; Score 820; DB 6; Length 2220;
Best Local Similarity 64.4%; Pred. No. 4,4e-58;
Matches 1356; Conservative 0; Mismatches 655; Indels 93; Gaps 5;
Qy 2 TGGCGACGCTTGTGATGACAACTACGACTGTTACCCCAACCTGTTCCAGTACATCG 61
Db 71 TGGCAACCTGTGATGACAACTACGACTGTTACCTTCAACCTTCCAGATGCTGG 130
Qy 62 GCGAGGCGACCGGGCAACCCCGCTGCTGTCGCCCAAGACGCGAGTGTCCGGCTGC 121
Db 131 CCGAGGTGAACGGCGCGCTCCGCTCGTCCGCAACGACACCCCGACCTGGCAGG 190
Qy 122 CCGTGCAG-----GACTTCGACGCGATCGTGTGTCGCGGGGCCCGGAGCCCGGACC 175
Db 191 CCTGGGCGCGGGGCGAATTGACAACTGTCGTCTTACCCCGGCGCGGCGACCCCGCCA 250
Qy 176 GGGAGCGGACTTCGGAATCAGCCCGGGCGATCAACGACGCGCTCCGCTCTCG 235
Db 251 CCGACACCGACCTGGGCGCTCAGCGCGGGGTATCAAGATGGGACCTCCCGTGTGCG 310
Qy 236 GCGTTCGCTTCGCGCACAGGCGCATCGCCAGCTCTTTCGGCGGAACCTGTGCGTCC 295
Db 311 GGGTGTGCTGTGGGCAACAGGCTGTGCTGCTGCGCGGCGCGCGCTGTCCAGCAC 370
Qy 296 CCGAACCCATGACAGCGCGGGGTCTCGAGGTGCGGCGACACCGGCGAGAGAGCTTTCGGG 355
Db 371 CCGAACCTTTTCAAGCGCGGACAGCGACATCCGCCACGACGGGCGAGGCGCTGTTCGCA 430
Qy 356 GCGTTCGCTTCGCGCACAGCGCGGTGCGTACCACTCCCTGCGCGCACCGACCTCCCG 415
Db 431 ACATCCCTCCCGGCTGACCGGTGTCCGCTAACCACTGTCGACCGTCCGCGCAACTGCGCG 490

Qy 416 ACGAGCTGAACCCCTCGCCTTGAGGAGCAACCGAGGTCGTCATGAGGCTTCGCGACCGCG 475
Db 491 CCGACTTCGCGGCGCACCGGCCACACCGCGACCGGCGAGCTGATGCGCTTCGCCACCGCC 550
Qy 476 AGAAGCGCTGTGGGGGCTTCAGTTTCACCCCGAGTTCATCGGACGACATTCGCGCGG 535
Db 551 ACTGCGCGGCTTCGCGGTGCAATTTCACCCCGAATGATCAGACAGGAACGCGGCACC 610
Qy 536 AGATCATGCGCAACTTCGCGGACCT----- 560
Db 611 GGAATGCTGCGCAACTTCGCGGACCTTCCTGCGCGCGCGCGCGCACCGGCCCGCAC 670
Qy 561 -----GCGCTGCGCCACACCGAGGACCGGCGGCGCACCGGGCGCATCTCC- 604
Db 671 CCGAAGCATATCCCGACCGCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
Qy 605 -----CGTAGAATCCACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646
Db 731 CGTCCGCGCGCGGAGGAGTACCGGCTGATGTCGCGAGTTCGCTGCGTCCCGCGACG 790
Qy 647 CCGAAGAGTATCCCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 706
Db 791 CCGACCGCGGTTTCACTGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 850
Qy 707 GCTCCGTCCTGGAAGCGCTGCGCGCTTCTTCTTCGCGCGCGCGCGCGCGCGCGCGCG 766
Db 851 GCGCGTTCAGACCGGCGCTGCGCGCTTCTTCTTCGCGCGCGCGCGCGCGCGCGCGCG 910
Qy 767 CCGAGTACCTCACTTACCGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
Db 911 GCGAAGATCACTTACGAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
Qy 827 CACGACCGGAGCGCGCGCGCTTCTTCACTTCACTTGAAGAGACGCTCGAACCGCGAC 886
Db 971 GCGAGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
Qy 887 GGGTCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
Db 1028 GCGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
Qy 947 GCTACGAGCTGAAGGCGGAGACCAACCGGACCCCGCGACCGTTCGCGCGCGCGCGCG 1006
Db 1088 GCTACGAGACCAAGGCGGACGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1147
Qy 1007 CCGCTTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066
Db 1148 GCGCTTCAATGTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1207
Qy 1067 TGTGCGCTTCGACCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1126
Db 1208 TCTTGACACTGAGACGACCGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1252
Qy 1127 CCGAGACCCCTCACCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
Db 1253 CCGAAGCTGTGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1312
Qy 1187 TCGGATTCGCCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1246
Db 1313 TCACCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1372
Qy 1247 AGCGCTTCAAGGCGCATTCGAGAGTGTCTCAAGAGATTCGCGAACCGGCGAGTGTAG 1306
Db 1373 CCGCTAACCGGGAATGCTGAGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432
Qy 1307 AGATTCGCTGACCAACATGATTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1366
Db 1433 AGGTGTGCTGAGCAATGCTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTCGAGTTC 1492
Qy 1367 CCGCGTGTGCGCGCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1426
Db 1493 GCGCCTGTGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1552
Qy 1427 CCGTGTGAGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486

Db	1553	CCGTGCTCAGCTCTCTCACCCGACGGTTCTCTCCGATCGGCGCGGACCGATTGGCGGAGT	1612
QY	1487	CCAAAGCCCATCAAGGGGGAACCCGCCCCCGGGGGGAGCACCGCGAGGAGAGACGACGGCTCC	1546
Db	1613	CCAAAGCCCATCAAGGGGCAACCCGCCCGGGGGGCGGGCCCGCCAGAGACGGCCGTC	1672
QY	1547	GGCGCCGACTCTGGCCCGCGGGGAGAGACCGGGCCGAGAACTGTATTCGTGCACCTGG	1606
Db	1673	AGGCTCTCCCTCGCCCGCGGCGGAGAAAGACCGGAGGAAACCTGTATGTCGTGCACCTGG	1732
QY	1607	TCCGCAACGACCTCAACAGACGCTCTCGCGGATCGAGCTCCGTCACAGTCGCCCGGCTTCGG	1666
Db	1733	TCCGCAACGACCTCGAGCCAGGCTCTCGACATTCGCTCCGTCACATACCGGACCTGTTCC	1792
QY	1667	AGGTGAGACCTTACCGCGCCCGCTGCACCACTGTGTGCATCCGATCCGGGAGACGGCTCGGC	1726
Db	1793	AGGTGAGACCTTACCGCCACCGCTGCACCAAGCTGTGACGACCGGTCCGGCCCGCTCGGGG	1852
QY	1727	CCGGACACGACACCGCCCGCTCGTACGCGCGCCCTTCCCGGCGCTTCATGACCGGCG	1786
Db	1853	CCGAGCTCTCCCGCCCGCGCGGTGACGGGCGGCTTCGCCGGCGGTGTGATGACCGGCG	1912
QY	1787	CGCCCAAGAGCGCAACATGGAGATCATCGACCGGCTGAGAGAGAGCCCGCGGGCGCT	1846
Db	1913	CGCCCAAGGTCCGCAACATGACGTTTCATCGACCGGCTGAGAGAGAGCCCGCGCGCTGT	1972
QY	1847	ACTCCGGGGCGCTCGGATGTTTCGCCCTCAGCGGGCGCCGACCTCGACATCGTCAATCC	1906
Db	1973	ACTCGGGCGCGTGGGCTTACCTTCGCCCTCAGCGGGCGCGGACCTCGACATCGTCAATCC	2032
QY	1907	GCACCATGCTGCTGGCCCGACGCGCCAGCGGAGGAGTTGGCGTGGCGGGCGCATGCTGTCC	1966
Db	2033	GCACCATGCTGCTGCCACCGAGGAGGCGCGCACCATTCGGGTGGCGGGCCGCTGTGTGCC	2092
QY	1967	TCTCCGACGAGGAGGAGGAGTTCAACGACGCGGTGTAAAGGCCCGCGCCANTGGTCAACCG	2026
Db	2093	TGTCGACACCCGACGACGAGGTTCCGCGAAATGCTCTCTTAAGGCGCAGACACACCTCGCGG	2152
QY	2027	CCCT 2030	
Db	2153	CCCT 2156	
RESULT 7			
SPU60417		4740 bp	DNA linear BCT 07-MAR-1997
LOCUS			
DEFINITION			Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine precursor biosynthesis (papa, papc, papb, papm) genes, complete cds.
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCES			
AUTHORS			
TITLE			
JOURNAL			

FEATURES	France
source	Location/Qualifiers
gene	1. .4740
CDS	/organism="Streptomyces pristinaespiralis"
	/mol_type="genomic DNA"
	/db_xref="taxon:38300"
	68_ .2227
	/gene="papa"
	68_ .2227
	/gene="papa"
	/function="p-aminobenzoate synthase"
	/codon_start=1
	/transl_table=11
	/product="papa"
	/protein_id="AAC44866.1"
	/db_xref="GI:1575336"
	/translation="MRTVPTLILIDNYDSPTYNLFWLAEVNGAALPVVANDPTRTWQALAGDPDNNVYVSPGHPNATDIDLSKRVITBWDPLIGVCLGHQALCLLGAAYHAPPEFGRTSIDIRHDOQLFANIPSLTVYHSLTVRGRPLVDLRTAHTAAGQMLAVAHNHLPRFGVDFHESISEBHGRLMANTRDLRLAAGRPPTTERIPAPAPAPAPAPAPAPGVEYRLVREYACVPDADAFALFADAPFMLDSRVPGIARFELGAAPGGEOTITVDADRAVRVYKDSGGETRRGTJEDLHEHLARALPAGLPEFNGVYVIGETYEKADSGDARHGEIPDGAFFPADMTALDHEOGAMTLASSTPRATPAARFMTDPAATLATTARRPPTLLPDQPLALDVYHRSLEPRYELVBECCRLITDGEYEVCLTNMLRVPGRIDPLIYRALRTVSPAPVAYIQFPGATLVSSPEEFLRIDADMAESKPIKGTTRPGAGAPADAAKASILAEXKORSENMLVIDVRNDLGVCDISAVHFGILFEVEIATYQLOVSTVRGLADVSPRAVRAAFPGSGMTGAPKVTMCFIDRLKEKRGVYSGALGYFALSGAADIIVRTIYATEAATIGVGAAVVALSDPDDEFRMLAKQTLTALROAHAGATSDRELLAGSLR"
gene	complement (2486. .3376)
CDS	/gene="papC"
	complement (2486. .3376)
	/gene="papC"
	/function="dehydrogenase"
	/codon_start=1
	/transl_table=11
	/product="PapC"
	/protein_id="AAC44867.1"
	/db_xref="GI:1575337"
	/translation="MRGSGVFCRCVVVGAGAVGRMFSHLVRSQVAVTWLVAAGAADVRVAGDVRRPGPEAVLAALADYVVLAVPEPAMEAVELVAGVMRGVILADTLSTKSRILAGRLREAPSLQAVGLNPMRPAEPLGIGQGRVAAVYTTDGGVATVELVAGWGAIVEMPARHDELTAQQAATHAIVLFGILGELSDVALKDSAPFAPLMEALGLAIAGTPEYVFEDIQANGPAPARQALGRGLVRLGQAVERGETFALPRLRGVLCGHGAELERLCARMPFALH"
gene	3413. .3802
CDS	/gene="pabB"
	3413. .3802
	/gene="pabB"
	/function="mutase"
	/codon_start=1
	/transl_table=11
	/product="PabB"
	/protein_id="AAC44868.1"
	/db_xref="GI:1575338"
	/translation="MTAAPITPAAPRATGPAATDPLDALRRLIDADAALLDAVTRTLDILRIKGEYKRLHQVPMQPHRIQAVHANAAHYAADHGIDPAFLRTLYDTITTCRLLEDWIASGGAPVPTPVHASASARGAVS"
gene	3799. .4677
CDS	/gene="pam"
	3799. .4677
	/gene="pam"
	/function="N-methylase"
	/codon_start=1
	/transl_table=11
	/product="Pam"
	/protein_id="AAC44869.1"
	/db_xref="GI:1575339"
	/translation="MTAAAPTLLAOLDEATGQTGAGITADAPADATRLRLAAACOVA PGDLDTCLAGVPPEPRFMHYVRRRLREAPERLVGHAIVYGGHFDLADGVFVKKPTEEITRDIAIRLEBALYRRTTAPLVYDLCAGGTMAVTLARVAPARVGLIELSQAARARARAKRTGIRIVOGDARDAFPELISGVLDVVTNPPITPLGLRTSAEVEVHEDPILAMAGEEGIMRAMERTDAARLAPGATLLHEHSGYOLASVAPLPRATGRSGHASPICN

ORIGIN

DGLJAVRNHTCAPPA"

Query Match 39.8%; Score 820; DB 1; Length 4740;

Best Local Similarity 64.4%; Pred. No. 3,2e-58; Matches 1356; Conservative 0; Mismatches 655; Indels 93; Gaps 5;

QY 2 TGCGCAGCTTCTGATGACAACTACGACTGTTACCCACAACTGTTCCAGAACATCG 61
DB TGGCAACCTCTGCTATGACAACTACGACTGTTACCTACAACTCTTCCAGATGCTGG 137
QY 62 GCGAGGCCACCGGCAACCCCGCTGCTGTCGCCAAGACGCCAGCTGTGCGCGTGC 121
DB 138 CCGAGGTGAACGGCGCGCTCCGCTCGTCCGCAAGACGACACCCGACCTGGAGG 197
QY 122 CCGTCGAG-----GACTTGAACCGGATCGTCGTCTCCCGGAGCCCGGACGCCGAC 175
DB 198 CCGTGGGCGCGGCGGACCTTGCACAACTGCTGTCTACACCGGCGCCGCGCACCCGCA 257
QY 176 GGGAAACGGGACTTTCGAAATCAGCCGCGGCGGATCACGCAACAGGCGGCTGCGCTCG 235
DB 258 CCGACACCGACCTTGGGCTTACGCGCGCGGATGATCACGAATGGAGACTTGCCTGCTCG 317
QY 236 GCGTCTGCTTCCGCGCACAGGAGCATCGCCAGCTCTTCCGCGGAAACCGTCCGCTCCCG 295
DB 318 GGGGTGCTGGGCGCACAGGCGCTGTGCTGCTGCGCGGCGCGCGCTGTCCAGCGAC 377
QY 296 CCGAACCCATGACAGGCGCGGCTCTCCAGGTGCGGACACCGCGAGAGAGCTTCCGGG 355
DB 378 CCGAACCTTTCACAGGCGCGGACACGACGACATCCGCGACGAGGCGGAGGCGCTTTCGGA 437
QY 356 GCGTCCCTCGCGCGCTTACCGCGCGTGGCGTACCATCCCTGGCGCGGACCGACCTCCCG 415
DB 438 ACATCCCTCTCCCTCTACCGCTGAGTCCGCTTACCATGCTGACCTGCGGACCTGCGCG 497
QY 416 ACGAGCTCGAACCCCTTCGTCGACGACGACGCGGCTGTATGGGCTTCGCGACCGCG 475
DB 498 CCGACCTGGCGCGCGCGCGCGACCGCGCGCGCGCGCGAGTGGCGCTGCGCGACCGCG 557
QY 476 AGAAGCGCGCTGTGGGGGCTCCAGTTCCACCGCGAGTCCATGCGGACGACTTGGCGGG 535
DB 558 ACCTGCGCGCTTTCGCGGTGCTGATTCACCCCGGATGATGACGAGAACACGCGCACCC 617
QY 536 AGATCATGGCCAACTTCCGCGACT----- 560
DB 618 GGAATGCTCGCAACTTTCGCGACTTTCCTTCGCGCGCGCGCGCACCGCGCGCGCAC 677
QY 561 -----CGCCTCGGCCAACACCGGGACCGGCGCGCACCGGGCGGACTTCC- 604
DB 678 CCGAACCGATACCGCGACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 737
QY 605 -----CGTACGAATTCACAGTTCGCGCGCGCTGACGTCGTCGCGACG 646
DB 738 CGTCCGCGCGGTGGGAGATACCGGCTGATGTCGCGAGTTCGCTGCGTCCCGACG 797
QY 647 CCGAAGAGTACCGCGCGGCTGCGCGCGGCGGCGGACACAGTTCGTGCTGAGACGCA 706
DB 798 CGAGCGCGGCTTACCGCTGTTTCGCGACCGCGCGCGCGCGCGGCTTTCGCTGACGCA 857
QY 707 GCTTCGCTCTGAGAGCGCTTCGCGCTTCTCTTCGCGGACGACCGCGCGCGCTCG 766
DB 858 GCGCGCTGAGAGCGGCGCTTCGCGCTTCACTTTCCTGCGCGCGCGCGCGCGCGCTCG 917
QY 767 CCGAGTACCTACCTTACCGGTCGCGAGCGGCTGCTTCCTGCGCGGCTTCGAGCGCA 826
DB 918 GCGAAGATACCTTACGAGTTCGCGCGGCGCGGCGCGTTCGAGAGAGCGTTACGCG 977
QY 827 CCAAGACCGGAGCGCGCGCTTCTTCACTTACCTGAGAGAGAGTTCGAAACGCGAC 886
DB 978 GCGAGACCGCGCGCGCGCGCGCTTTCGACACCTGAGAACAGAACTG---CGCGCC 1034
QY 887 GGGTCCCGCTGCGCGCGGACGCTGCTTCGAGTTCAACTCGGCTACGCTACGCTCG 946

DB 1035 GCGCGCTGCGCGCGCACCGGCGCTGCGCTTTCGAGTTCAACTCGGCTACGTCGCTCG 1094
QY 947 GCTACGAGCTGAAGGGGAGAACACACCGGACACCCCGCGACCGGCTCCCGCACCGGACG 1006
DB 1095 GCTACGAGAACGAGCGGACGAGCGGCGCGAGAGACCGCCCGCGGAACTCCCGACG 1154
QY 1007 CCGGCTTCTCTTCGCGGACCGGCGCATGCGCTTCGACGACGAGAAAGCTGCGTACG 1066
DB 1155 GCGCTTATGTTTCGCGGACCGGATGCTGCGCTTCGACGACGAAACGAGGCGGCGCTGCG 1214
QY 1067 TGTGCGCTTCGACCGCGGAGCGACGACGAGGCGCGCGGCTGCGTTCGAGAGACG 1126
DB 1215 TCTGGGACTGACGACGACCGGACGAGCGCGCGCACCGGACCGCG-----G 1259
QY 1127 CCGAGACCTTACCGGCTGAGCGCTTCGCGCGCGCGCGCGGACCGGACCTCCGCTGCT 1186
DB 1260 CCGAAGCTGAGCTTACCGAGCGCGCGCGGACCTTCGACACACCGCGCGCGCGCGCT 1319
QY 1187 TCGGATCCCGGAGGCGCGCGCGGCGGCTTTCGCGCGCGCGCGCGCGCGCGCGCGACG 1246
DB 1320 TCACTGCTGCGCGGACGACCACTGCGCGCGCTTGAAGTTCATACCGCGACGCTTCG 1379
QY 1247 ACGCTTACCTTCAAGCGCATGACGAGTGCCTCAAGAGATCCGCAACGCGAGTGTACG 1306
DB 1380 CCGGCTACCGGAGAACTGTGAGAGAAATGCGCGCGCGCGGATTCACGACGCGGAGACCT 1439
QY 1307 AGATTCCTGACCAACTGTGTACCGCGCGGACCGGCGGAGCGCGCTTCGCTTACT 1366
DB 1440 AGGTGTGCTGAGAGAACTGTCTCGGAGTGCCTGCGCGCGGATTCGACCGCGCTTACC 1499
QY 1367 CCGCGCGCGCGCGCATGACCGCGCGTCCGTAAGGCGCGCGCTGCGAGTTCCCGGAACTGT 1426
DB 1500 GCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1559
QY 1427 CCGTGTGAGCGCGCTGCGCGCGGAGCTTCTCAAGTTCGCGCGCGGACGCGCGCGTGA 1486
DB 1560 CCGTGTGAGCTTCTTCAACCGGAGGTTCTGCGGACATTCGCGCGGAGCGGTTGGCGGAG 1619
QY 1487 CCAAGCCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1546
DB 1620 CCAAGCCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 1679
QY 1547 GCGCGACCGTGGCGCGCGCGGAGAGACCGGCGCGGACCGGATGATGCTGACCTGG 1606
DB 1680 AGGCTTCCTTCGCGCGCGCGGAGAGACCGGACGAGAACCTGATGATGCTGACCTGG 1739
QY 1607 TCCGAAACGACCTTCAACAGCGTCTGCGCGCTTCGCTTCAAGTTCGCGCGCGCTTTCG 1666
DB 1740 TCCGAAACGACCTTCGCGCGGAGTTCGCGACATGCGCTCCGTCACGTAACGGGCTGTTCG 1799
QY 1667 AGGTGAGACTTACGCGCGCGGTCACAGCTGTGTGACATCCGGGACGCGCTCGCG 1726
DB 1800 AGGTGAGACTTACGCGCACCGTTCACAGCTGTGACAGCGTTCGCGCGCGCTCGCG 1859
QY 1727 CCGGACCAAGCACCGCGCGCTGCTGACGCGCGCTTTCGCGCGCGCTTCATGACCGGG 1786
DB 1860 CCGAGCTTTCGCGCGCGCGGAGTACGAGCGCGCTTTCGCGCGGCTGATGACCGGG 1919
QY 1787 CGCCCAAGAGCGCACCATGAGATCATGACGCGCTTGAAGAGAGCGCGCGGCGGCT 1846
DB 1920 CGCCCAAGAGCTCGGACCATGAGTTCATGACCGGCTTCGAGAGAGCGCGCGGCGGCT 1979
QY 1847 ACTCGGAGGCTCGGATGATGCTTCGCTGAGCGGCGCGCGGACCTTACGATGCTATCC 1906
DB 1980 ACTCGGAGGCTCGGATGATGCTTCGCTGAGCGGCGCGCGGACCTTACGATGCTATCC 2039
QY 1907 GCAACATGATGCTGCGGACCGGCGGAGGATTCGCGCGGCGGAGTTCGCTCC 1966
DB 2040 GCAACATGCTGCGCACCGGAGGCGCGGACCATTCGCTGAGGCGCGCGCTGCTGCC 2099
QY 1967 TCTTCGACGAGAGAGAGTTCACCGAGACCGTGTAAAGCGCGCATGTGTCACCG 2026
DB 2100 TGTTCGACCGCGACGAGAGTTCGCGAATGCTCTCAAGGCGGAGACCACTTCGCG 2159

Db 2267 ACCGTGAGAGCGCTTCTGACCTGCGTAGAGAACGACCTGGGGGCTGCG---CACG 2323
 QY 895 GTGGCCCCGACATCGCCCTTGAAGTTCACTCGGCTACGTCGGCTACCTCGGTAGAG 954
 Db 2324 GAGGTGCCCCAATCTCTTCTGCGCTTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 2383
 QY 955 CTGAAGCGGAGACACACCGGCGACCCCGCGACCGGCTCCCGGACCCCGGACCGCGCTTC 1014
 Db 2384 TTGAAGCGGAGACACACCGGCGACCCCGCGACCGGCTCCCGGACCCCGGACCGCGCTTC 2443
 QY 1015 CTCTTGGCGGACCGGCGCATCGCCCTGACACACGAGAGAGCTGCTTACCTGCTGGCC 1074
 Db 2444 GTCTTGGCGGACCGGCGCTGCTGCTGACACACCGGACCCCGGACCACTTACCTGCTGGCC 2503
 QY 1075 CTGAAGCGGCGGCGGCGACACGAGCGCGCGCTGCTGGCTGGCGGAGACCGCGGAGACC 1134
 Db 2504 CTGATGAGAGACACACCGGCGGCGGCGGCGCTGCTGGCGGCGGCGCTGCTGGCGGAGACC 2563
 QY 1135 CTGACCGGCGCTGGCGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1194
 Db 2564 CTGACCGGCGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2599
 QY 1195 CCGGAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1254
 Db 2600 CCGGAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2659
 QY 1255 CTGAAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1314
 Db 2660 CTGAAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2719
 QY 1315 CTGAAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1374
 Db 2720 CTGAAGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2779
 QY 1375 CCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1434
 Db 2780 CCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2839
 QY 1435 AGCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1494
 Db 2840 AGCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2899
 QY 1495 ATCAAGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1554
 Db 2900 ATCAAGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2959
 QY 1555 CTGAGCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1614
 Db 2960 CTGAGCGGCGGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3019
 QY 1615 GACCTCAACAGCGTCTGCGGCGGCGGCTTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1674
 Db 3020 GACCTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3079
 QY 1675 ACCTAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1734
 Db 3080 ACCTAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3139
 QY 1735 AGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1794
 Db 3140 AGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3199
 QY 1795 AAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1854
 Db 3200 AAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3259
 QY 1855 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1914
 Db 3260 GCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3319
 QY 1915 GTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1974

Db 3320 GTGCTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3379
 QY 1975 CAGGAGAGAGAGTTCACCGGACCGGTAAGAGCCCGGCGCATGCTACCGGCTGAC 2034
 Db 3380 CCGGCGGAGAGTTCACCGGACCGGTAAGAGCCCGGCGCATGCTACCGGCTGAC 3439
 QY 2035 GCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2057
 Db 3440 ACCGCTTCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3462
 RESULT 9
 SGR300302
 LOCUS
 DEFINITION
 STREPTOMYCES GRISEUS PARTIAL ORF1, CANA GENE, CANC GENE, CANF GENE,
 cant gene, cana gene and canR gene.
 ACCESSION
 AJ300302
 VERSION
 AJ300302.1 GI:12231153
 KEYWORDS
 ABC-transporter; aminotransferase; can RB gene; cana gene; canc
 gene; canf gene; canR gene; cant gene; Cho-like protein;
 cytochrome P-450; ferredoxine; PABA synthase; thioesterase.
 SOURCE
 Streptomyces griseus
 Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 Streptomycinae; Streptomycetaceae; Streptomyces.
 ORGANISM
 Streptomyces griseus
 REFERENCE
 1 Campelo, A.B. and Gil, J.A. Microbiologia, Universidad de Leon,
 Thesis (2000) Department of Microbiologia, Universidad de Leon,
 Leon, Spain
 2 Campelo, A.B. and Gil, J.A. The candidicin gene cluster from Streptomyces griseus IMRU 3570
 Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)
 MEDLINE
 21642576
 PUBMED
 11782498
 REFERENCE
 3 (bases 1 to 39314)
 AUTHOR
 Gil, J.A.
 TITLE
 Direct Submission
 Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de
 Leon, Campus de Vegazana s/n, 24071, SPAIN
 JOURNAL
 Location/Qualifiers
 1..39314
 /organism="Streptomyces griseus"
 /mol_type="genomic DNA"
 /strain="IMRU 3570"
 /db_xref="taxon:1911"
 complement(<1..1785)
 /note="ORF1"
 /codon_start=1
 /craen1_table=11
 /product="Cho-like protein"
 /protein_id="CAC22112.1"
 /db_xref="GI:12231154"
 /db_xref="UniProt/TREMBL:O9EWC6"
 /translation="MLERTETIQLVDAALGATBRSSVLVTEPLGIGRSALLORL
 SCIADRDVVRVLAADAPMEDPFAFVGVRQLPETLLGSPEDARDMDADASFARH
 VIADDAAPGADVAIAATEAVDGLSLANYSABSRLLIVDDLOWSPVPSRWLT
 LKRLHGLAAVVVCAALRDPSRSHLTVEIRBAATQTLRPASLSLAHELVREHFG
 EAGDDEFVQCHGASVGNPLFLSLIVGSPFGRPLAETARLRPSOURERLAS
 ILRTQAPVPRDLAAALAILGSDAPTLARLAGLDSIGVAGALRALGALAAAPDP
 RFIHRSVRDAESTLTMVOREKMDDEAALLIYAAGSPAQOVAQMLGAVYTRQPAVD
 VIRAADVTALRRGAPDTAGVLCRALDSPAAGVGGRLLVELGTABERFPDPLACERH
 IAQAMTLRPEPRDRAVALRISPTALGPALVAVLRLQAAADLGPADGLTGRDLA
 LLEARLRLPCHGEDPHELAASARLKEIGPEPVPDTGGERELTIALGAATLGCHLPA
 SEVVRAGVLRREPATSARISHTLPLAI"
 9525..10583
 /gene="cana"
 /gene="cana"
 /gene="cana"
 9525..10583
 /codon_start=1
 /craen1_table=11
 /product="aminotransferase"
 /protein_id="CAC22113.1"

```
/db_xref="GI:12231155"
/db_xref="GOA:O9EMC5"
/translacion="MAFTHPVSRPALDGRDELTVYVSDAVSGGWSISOGPYVVRPEEPA
EMNGVAGVACSSGTAALTLALNINIGREDLTVYVPTVNASAMVTTGATPVVVD
CGDDUNIDVTRIEKITTPTTRAVMPVHVGRCDMDVMDLQVNLRVEDSEAHG
VRVVDIACPSLPAKKITVGEQVCLTDDPRLAQHILRAMAFTRDSFHLKJLY
NYMTMAGVVALAQTERLDEILATREIARVDAGIKDIPGITTLPARSDVLMYDI
AERREELRAHLDERGIETRLFFKPMRQPGYLDPPVPTINAHFSESDGLYPHTGT
AADQYITGAVDPFYRAG"
10678..11859
/gene="canC"
10678..11859
/gene="canC"
/codon_start=1
/transl_table=11
/product="cytochrome P-450"
/protein_id="CAC22114.1"
/db_xref="GI:12231156"
/db_xref="GOA:O9EMC4"
/db_xref="UniProt/TREMBL:O9EMC4"
/translacion="MTTSPGTVVDPPTPTPTREPLPLSOYAEHRKONGIYQTHLNGR
PIVLVTRHEDYRAVLTHPRI SANPDNBPNGVETNGVPOEOI PGFVGLDSPEHR
ERKVLPEFTVRVRELRPAIERTVBERIDAMLAGNTADLVNDFALPVSIVISALL
GVPSADRPFPESRTKLVAIRISIDBERAATROLRYINRLVIKKKRGEDLISRL
LSTGKLSDELSGVLLILLIAGHETTANNIGLVVTLISHEWIGDDRLVEEDLRLS
VADVMALRVAVDVDEIAGOTIRKGEIGVPLASANDTEAFGCPHAFNPRTERRHVA
FGYGVHQCIGQNLVVRVMEIAYRKLPERIDELALVPEOQAKYDGIIEGLHELEVR
W"
11906..12100
/gene="canF"
11906..12100
/gene="canF"
/codon_start=1
/transl_table=11
/product="Ferredoxine"
/protein_id="CAC22115.1"
/db_xref="GI:12231157"
/db_xref="GOA:O9EMC3"
/db_xref="UniProt/TREMBL:O9EMC3"
/translacion="MRYTVDSEQCVAGGCVLNAPVFPDODDGVVLLRAEPDERDH
EAVRIAGDLCPSSAVVQBD"
12223..12993
/gene="canT"
12223..12993
/gene="canT"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="CAC22116.1"
/db_xref="GI:12231158"
/db_xref="GOA:O9EMC2"
/db_xref="UniProt/TREMBL:O9EMC2"
/translacion="MTTADDTAARMLRHRHPADA VRLVCFPHAGSASFYHPVSAR
FAPGAEVVSLOYPGRQDRKEPCVPDLGTADLITBOLLPLDERPTVFGHSGALIA
FETAMRLBQKAGPRTVIASGRGSPSTTAEVHTDDGIVEMKRLNLTAGVAGALA
BEILRALPLRLRGDYRAIETVTCPPDRILACGLTVLTGEDDPLTTVEAEHRMDHTG
PFLRLVFTGGHFLITQLDAVNTIEINQALHPDAARA"
13190..15361
/gene="pabAB"
13190..15361
/gene="pabAB"
/codon_start=1
/transl_table=11
/product="PABA synthase"
/protein_id="CAC22117.1"
/db_xref="GI:12231159"
/db_xref="GOA:O9EMC1"
/db_xref="UniProt/TREMBL:O9EMC1"
/translacion="MRTLLVDVNDSTFYNLPHYLSRANGREPVYINDDPANRGLLD
APFNVYLSPEGTGPHRPADGICARIAEBRILPVLCVCIHQOMALAHARVGRAPED
RHGRTSAVRHDGTGLEGLPQPLEVVRHYSLAVTELPELEAHGMALEDVIALARIT
LPLMGVQFHPESIGTQDGHRLANFRDLTERHGRTRHGRAGHTLPPAPARETKAT
```

```
TCGPRRLRYIAKSLPTRMDEAVAFDSLFRITGDHPWMLDSRREGELGQLSMMGDASGP
LARIKADHAGTAVTRADGASTVESAPLTVLWEDLADIARTEVELPPAPALGMWGC
LYELKABEDGDAHARSPPDAVLVPADBALVLDHRTTLYLAREVEDDAEABAMWL
LAASATDNLVADREBPCEAPVCTGPELIRHDDGYKLDVCOQETIACTEYEVIC
LTMABADDTLTPMAIYRRLRVSPAPFAFLIDPGMAVLSSEPFLITDHGRMES
KPIKTRPBGATPOBDALVRLATCEKRAENLMI VDLVRHDLRCAEVGSVADPV
PVEYVATVHQVLSVTYLRIDREDSPVAVRAAFPGSGMTGAPKRTMQIIDRLGCP
RQVYGALEYFELTGAVDLSIVIRTVYSGGLRYGVGGAIVIALSDPADEFRETVKA
AFLRLDLPAGPREAPGKDLDBEPDDGDGADAPDVLVPG"
20829..21836
/gene="canRA"
20829..21836
/gene="canRA"
/codon_start=1
/transl_table=11
/product="ABC-transporter"
/protein_id="CAC22118.1"
/db_xref="GI:12231160"
/db_xref="GOA:O9EMC0"
/db_xref="UniProt/TREMBL:O9EMC0"
/translacion="MNGIHLPPAPLPAQVSAAGSGDLMTEAGVSKAYGTHALDSV
SLAUVRGTYGLLHNGAGKTTVLVILITTALEPTTGRARVAGYDADRPVEIRRRIGL
TQCFASVDAQLSGYDNVLVILALGAGRAARVRARELLELFRLEVRADRASSYGG
LRRRLDLAVSLVGAPEVLFLDEPTGLDPSRINIMEIVELVBEQTVVLLTOYLEE
ADRLDLAVLSAGVVAAGTAPELKATVGRVTVLTLETGDSVAARATALGAFAP
VGEERATVVPIDATREIADIIRSLDQAGVEASELNFGEPTLDVYTLTLAGAQRTR
A"
21970..22776
/gene="canRB"
21970..22776
/gene="canRB"
/codon_start=1
/transl_table=11
/product="ABC-transporter"
/protein_id="CAC22119.1"
/db_xref="GI:12231161"
/db_xref="GOA:O9EMB9"
/db_xref="UniProt/TREMBL:O9EMB9"
/translacion="MPRAVAPVAPFGSSLFTQIRVLTGRSLRAMVTDPIVFLGIOP
VILVLVLDVRESKMGMPHFPPGVSYLDIVLPAVLVDNAQSMGSGVLVVDQKNGI
VARLSLSPHFGALIAARSLVGLVRSAAQVATMLAMTVLGSISNGAELTASAGL
TLFISFLGMAPIAAGAMLRAPLQNLALIVIFPLMPASSAVYVADLPSTLGAVAS
VNPLTVADITRALALDVPGLDHAIPAVVICAVIAVGVMAVRGFRBPL"
ORIGIN
Query Match 37.5%; Score 773; DB 1; Length 39314;
Best Local Similarity 63.3%; Pred. No. 8.5e-55;
Matches 1337; Conservative 0; Mismatches 680; Indels 96; Gaps 6;
QY 1 ATGCGAGCGCTTCTGATCGACAACGACTGTTCAACCACACCTGTCAGTACATC 60
DB 13190 ATGCGACCCCTTCTGTCGACAACTACGACTGTTCACTTACAACTCTTCACTACCTC 13249
QY 61 GCGAGAGCCACCGGAGCAACCCCGTCGTGTCGTCGCCAAGCAGC---CCGACTGTGTCGCG 117
DB 13250 TCCCGGCGCAACGCGCGGGAACCCGAGGTCATCCGAAAGCAGACCCGCGCTGGGGCGCG 13309
QY 118 CTGCCCGTGAAGACTTGAAGCGGATCGTCGTGTCCTCCCGGAGCCCGGACGCCGAGCGG 177
DB 13310 GGTCTGCTCGACGCGCTTGCACACGTCGTGCTCTCCCGGGGCGCGGACCCCGACCCG 13369
QY 178 GAACGGGACTTCGGAATGACCGCGGGGCAATCACCGACAGGGGCGTGGCTCCCTCGGCG 237
DB 13370 CCGGCGGACTTGGCGCTGTGCCCGGAGATCGCGAGAGAGGCGCGGCTGCGGTGCTCGGC 13429
QY 238 GTCTGCTCGCGCACCGAGGCAATCGCCAGCTCTTTCGCGGGAACCGTCGCGCTCCGCCCG 297
DB 13430 GTCTGCTCGCGCACCGAGGCAATGCGCCATCGGCCACCGGCGCGCGGTCGCGGCGCCCG 13469
QY 298 GAACCATGCAAGCGCGGCTCTTCCAGGTCGCGGCAACCGGCGAGAGAGCTCTTCCGCGGCG 357
DB 13490 GAGCCCGGTCACGCGCGGCACTTCGCGGTCGCGGACGACGCGGACCGGCGGCTCTTCAAGGGG 13549
```


Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China

FEATURES

source

Location/Qualifiers

1. 138203

/organism="Streptomyces sp. FR-008"

/mol_type="genomic DNA"

/db_xref="taxon:206662"

complement (574. .1950)

/gene="fscO"

complement (574. .1950)

/gene="fscO"

/note="putative PAD-dependent monoxygenase"

/codon_start=1

/transl_table=11

/product="FscO"

/protein_id="AA082549.1"

/db_xref="GI:34766436"

/translation="WPSTCPAPTSIGRGCPHRRKGTSGRRRIAVTPRCGVKRGH

QAKKRVSMRGSVAVVGSIAGCAALASRGARITVFERADRLRDGVIGV

HDAVYELRDAGVADDEMPWAPLNRVWSTVADGAEHRAIGTPPEFRAYNGLMS

ELRRVETETVYAGAKVEAEODADGVTRLADGPEREDLVIGADGVSVREMY

PGTAAVAGYIGMRGTS PDVSGLPDGLDAHNTTFPGCHMAVRIPDGSGHRLNVL

YTAARRIDGHHPLRTPTSLPRGLNAELTEHRAIYAHEFPFPMAKLSTAEITF

IQPTVDVPHYATDRNVLIGDASVAKPHLGSVVALQDATALDEAWVAGSWEKI

LEGYHARGPVGTAMVGLARMGSAQVEETPDWSAMQAEFDAMWQENNDRSRGF

GCHLSKR"

complement (2264. .3037)

/gene="pabc"

complement (2264. .3037)

/gene="pabc"

/note="putative 4-amino-4-deoxychorismate lyase"

/codon_start=1

/transl_table=11

/product="pabc"

/protein_id="AA082550.1"

/db_xref="GI:34766437"

/translation="MIELDEPAGPEALASIALTNXGFTLLVENGVRGLDLHER

LIRDCRTLPPADLDPRVKRLARRAPTDGRAIVRVFPALNLGINIADAPGILV

TSRAPDKPGLPDRSVVRRLDPEVSGVGLCTTLRROAORAGYDVLFTGPDGD

ILEGTTNVLNRGVEVWPGHSTVLAETROLIRGRHSRHTGRPRGRLSGGLRD

QRVGVAPVTGIDDRFPAHNSVTRIAETIYQALPGSP"

complement (3150. .3818)

/gene="fscR1"

complement (3150. .3818)

/gene="fscR1"

/note="putative transcriptional activator"

/codon_start=1

/transl_table=11

/product="FscR1"

/protein_id="AA082551.1"

/db_xref="GI:34766438"

/translation="MDPAPADPAGSPERAALVADGPRAGAVVACIDPSLTIOVNO

EPERRPGSSSELGSRFCDLVHVSQOPLMHQPARULDGRRRFATEVYAUOERTA

STLPLNALVARGGKTPDPAALLVVMAABEADADWAPRKLLSTIDRIKILLEGIA

GVSTIPLASRLYLSRQGEVYHVTGLRKLVPRKALVSRASWGLKGTWPKYVE

DFIK"

complement (4377. .7205)

/gene="fscR11"

complement (4377. .7205)

/gene="fscR11"

/note="putative transcriptional regulator"

/codon_start=1

/transl_table=11

/product="FscR11"

/protein_id="AA082552.1"

/db_xref="GI:34766439"

/translation="MLLEETTELALVDALGAATEGRSSLVLTGVLGSRNALQRL

SCIADDDVRLVRAHAPMEODFAFVGNQLFETLLGSDSPEDARDMDHASFARH

VLADAPPGADOLIAATEAVLIGLILANVSADSLITLVDLOMSDPSIRMLTY

LAKRIHGLRAVVCALADGDRSHHTLAREIRAGCTTAPASISLTATHEVREHNG

EAGDDEFVQACHEASVGNPLFLUSILVGTGFLGRPLAEHETARRLRPSQLERLAS

ILRTQPAVRDLAALA1LIGQSDAPLALAGDSIGAGALRALGALGALAAPPBP

RFIHRSVDAESTLTWVORERMHDEAALILYAAGSPAEQVAAQOLMAVTVTRPPWADP

gene

CDS

complement (7210. .10254)

/gene="fscR11"

complement (7210. .10254)

/gene="fscR11"

/note="putative transcriptional regulator"

/codon_start=1

/transl_table=11

/product="FscR11"

/protein_id="AA082553.1"

/db_xref="GI:34766440"

/translation="WYDHPPIELPIERDRELAALSAVIGELSGRPVAVTVTGP

GIGONLILMAAAYATDGLRILSAHATPAHEVYGVVAQILAEENRLAARLPLTD

EQGGCPGLNGLNVRDRPTLLVRDTQWLDPASIRMLBARLPRAPVALITSTT

GTALTRPENSVPPLTGLATTTELALPULSTGCTAVVIRARAGADGDAFTALAEAT

RGIPIAVHVDLDRFARAGSPRADRIDPLRALTAEVGDHATRLDLDPAVVALR

ALAVCGDLDPFLVCTIAGPHSVSESRILAAALASGLITLRDGHPRVQDPAVVALR

EMPAADRAELTVARAAGLQVAVADOGGIDLLILLRPVDPNAVVDLRRVVALRG

RDLAVAYLARALDEPLAEDRARIIPOLASPMATVPAAERLIGLIRAPRPGA

GIRARATDCLIGGDTRAARHLAGIDBAAPPEPRPGTCAVPAVPEABSPRT

HLVPHRPVATGTPGATPDGRSAAPDLGETRAADAPSPGEEAAGRTDAGPDDP

AAVERLETVLCRVASFLRHDDADLDVLPVPLPAVHNPAAVAGWAMEHGVAGVIRE

EVRRLARSAPFVPAATGRPLLVPRLLGCPALLITDDGAEHNLGALDSEHREARAI

SVAHILTVRAEHLIRHGRPDAAARDLAAQALPDRRLPLFLVYALSMITDLONG

HTEBARETAARPLPLAHESATTAQOLPRTGVLTATDDEPHQAREHFRACGWLHNG

CANPAVQPRSLAABAAHTLGTTEBARLVHEEVRILARKMCAAFLGRQQLSAYVTE

ENKVENLRAAVTLLSASARTVTRAIVLELAABEENRRKRAVBLGPAATVLPAP

PGPHVPSVAVAPPGPRPAVLTGRRKALPGHDAEVMHGLTVERBTAALAAQGLGN

BIATELAVTTRAVELRLSGVRLKLRGHEELRALVQAECS"

complement (10298. .13315)

/gene="fscR1V"

complement (10298. .13315)

/gene="fscR1V"

/note="putative transcriptional regulator"

/codon_start=1

/transl_table=11

/product="FscR1V"

/protein_id="AA082554.1"

/db_xref="GI:34766441"

/translation="MMLRWAGDRSRGAAPGRSGSAGPSLAPSARAESFVPHRSRAP

PAPPLVGRAHQDLVLRHADARAGRPRLVLLDGPAGIGKTALLAALDEGPLAGMT

TLYGSCRAVDVATGYGVRLFGGLGTLGRKGTSPLAGGRRALPALAAPGELDA

DGSQTSVIGLWYLAANMADGPVLYVDVLTMDDEBSLRMLDTRADGLPLTLV

AAHRTGTCTATPADLADLVHMLPNSLCGPIDITAEVAILTHSPFEODPLSPFGR

LAVTGTSPLEIVRLRLDLRSAGLGPDDTALAAQGVVAASVAGVLEHOPDWRO

VARAAVVLGDEBRTYLAAGRVSVHVEBAVILIDPAGIHGHLELSDILVRAVL

AVAGSGVAVLRRRAARLLSDGRSPEEITQCLIPGTGDDMAVSLRDAQAQEOBG

ACEAARVLEERYBAEAPKDPVLSRGKALTEADPAESVTLHEAHSITTDVRAAT

AVOGLTCLAAVOOSPDGARALTEALDAELGPREPRADRELRTLTSSALLIVSDE

KATLPDITLRTEGLTPQEDTPAQROQLAMLSVLSABEGGAEETVGRARRLRAPGV

PLGVMSLTPSLALSADENEAEEVLELSSGDTGAWTVYVALSTSLFRLNG

AVDDMAADQTALEILIGORNGDTATMPTTAASVYTERGDEALALDAIKRPLD

RFVWEYHWYLMARGRLAADGDLGALGVGSCGASMAAGLTNPVLAWMLETCLL

GEAGGEEAGRAAAGTTRLAERWGTBRALGVYALRGARAGTARAGALREAVALLAD

SPARSGHARGLLILGRALVEDAVAREGRPHLEAENAGLARCCCVALARARABELIAG

GMRREYASPLMDLTGTERTVAGLVASGNGENVBESLFTVTRVELHLTSYRKLVG

ARRGDLTEALREAGATARPAAERRPQHAARNP"

/gene="fscM1"

/note="glycosyltransferase responsible for mycosamine

attachment"

/codon_start=1

```

/translate=11
/product="Fscm1"
/protein_id="AA082555.1"
/db_xref="GI:34766442"
/translation="KDSAAKPIILFVSLPPSGLLNPLVLVAGELARGVDPDLFATDDEH
RREVEVALDVSFVSFASLGEVVPBELSAATWDEVEVREYQSSRFAPAHAIVAQSYRP
ALQARKYRELQVWDEVRPALMNVDCVAFGVDLALARGI PYLAMP PYASNLVTSN
PGASATPKPSF PVPNSGLPARKMSRQKMLTFLPKRTGLFHPDUALLRDPAALIRK
ELGIAPNNMYMGRTITRLTREVALVVEARMSRTHQPLKLPREOHLPEASGL
WIDQASVYVMSQGLDVAHPNVSFPHSGGNAHGVYFGKQVVRPLWVDFDVA
PMLRVESVPSQDLVLAHPNVSFPHSGGNAHGVYFGKQVVRPLWVDFDVA
RGRDFGILSLTLDKPHTVDPDVKLTRVSDPAFRTEARLQRAAGRAAADL
VTGLLPPAA"
14953.16011
/gene="fscm11"
14953.16011
/gene="fscm11"
/notes="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/translate=11
/product="Fscm11"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTHPVSRPALDGRELEXYDVAVSGNISQGPYVRRPEEAF
ENMGVHGVASCSGTALTLALBALNGDEVIYVEFTMVASAMVTTGTGATPVVD
CGDDLNDIYTRABEKITPRTRAMPVHYVGRCDMDVMDLALQYTLRVBDSAEHG
VAPVDIACFSLPANKITTAGSGVCLTDPPRLAEQDLARAFATRDSFLMKLALY
NVRMTMQGAVALAOERLDELATRRLEIARYADGLKDLPGITLMPADVLMVDLR
AERRELRALHADRGIETRLFFKPMRSRQGYLDPVPTLNARHFSBDGLYLPHITGLT
AAQEXYTGAVRDYPYAG"

```

Query Match 37.4%; Score 769.8; DB 1; Length 138203;

Best Local Similarity 63.2%; Pred. No. 9.2e-55;

Matches 1335; Conservative 0; Mismatches 682; Indels 96; Gaps 6;

```

QY 1 ATGCGACGCTTGTGATCGACACTAGACTCGTTGACCCACCAACCTGTTCAGTACATC 60
DB 18610 ATGCGACCTTTCGTGACAACTAGACTCGTTACCTACCAACTCTTTCACATCACTC 18669
QY 61 GGGGAGCCACCGGGGACACCCCGGTGTCGTGCGCAACGACG---CCGACTGATCGCG 117
DB 18670 TCCCGGGCCAAACGGCGGGGAAACCGAGATCATCCGCAACGACCCGGGCTGGCGGCC 18729
QY 118 CTGCGCGTGAAGACTTTCAGACGAGATCGTGTCTCCCGGGGCCCGGACGACCCGACCG 177
DB 18730 GGTCTGCTGACCGCTTTCGCAACGTGTGCTCTCCCGGGGCCGGGTACCCCAACCGG 18789
QY 178 GAACGGGACTTGGGAATCAGCCGCGGGGGATGACCGACAGCGGGCGCTGCGCTCGCG 237
DB 18790 CCGGCGGACTTTCGCGCTGCGCGCGCGGATCGCGAGAGGGCGGCTGCGCGTCTCGCG 18849
QY 238 GTCTGCGTGGCCACCGAGGATCGCCAGCTCTTTCGCGGAAACCGTTCGCGCTCGCGCG 297
DB 18850 GTCTGCGTGGCCACCGAGGCAATGCGCTTCGCCACGCGCGCGCGGGTGGCGCGGCC 18909
QY 298 GAACCATGACGCGCGGGGTCTTCAGAGTCCGCGACACCGCGGAGAGACGTCTTTCGCG 357
DB 18910 GAGCCCGCTACCGCGCGCACTTCGCGGATACGCGACAGCGGCACTTCGAGAGG 18969
QY 358 CTCCCGCTGCGCTTCAACCGCGCGGTGCTACCTCCCTGCGCGCGCCACCGACTCCCGCG 417
DB 18970 CTGCGCGACGCTGAGAGGTGGGTGATCACTCCCTGCGGTGAGGAACTGCGCGCG 19029
QY 418 GAGCTGAAACCCCTCGCTGAGACGACGAGGGGTGTCATGAGGCTTCGCGGACCGCGAG 477
DB 19030 GAGCTGAGAGCCACCGCTGATGAGAGAGCGGGGTCTGATGCGGTGCGCGCACCGGAG 19089
QY 478 AAGCGCTGTGGGGGCTTCAGTTCCACCGGAGTCCATTCGCGGACGCACTTCGCGCGGAG 537
DB 19090 CTGCGCGCTGTGGGGGTGCGATTTCACCCGAGTTCATTCGCGCACCGAGCGGCTCACCG 19149
QY 538 ATCATGGCAACTTCGCGGACCTCGCGCTCGCCACACCGGGGACGAGC-----GCGAC 591

```

```

DB 19150 CTGCTGGCAACTTCGCGGACCTTCACGAGCGCGACGCGCGGACCGCGCGCGCGCG 19209
QY 592 GGGGCGGACTCCCGCGTACGAATCCACGCTGCGCGCGGTGAGAGTGTCTGCGGACCGCAA 651
DB 19210 GCGGGGCAAGGACGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 19269
QY 652 GAGGTACGCGCGCGGCTGCGCGCGCGCGAG----- 681
DB 19270 ACCGCAAGCGGCTCGGGTCACTCGCGAGTGTCTCCCAAGCGGCTGGAGCCGAGGTC 19329
QY 682 -----GCGACCACTTTCGCTGAGACAGACCTTCGTC 714
DB 19330 GCGTTGACCTGCTGTTTCGCAACCGGCGACACACCTTCGCGGTGACAGACGCGTCC 19389
QY 715 CTGGAAGCGGCTCGCGCTTCTCTTCCTTCGCGGACGACCGCGGCGCGCTCGCGAGTAC 774
DB 19390 GGGGCGGACTGGGCGGACGCTTCGATGAGGCGAGCGCTCAAGTCTCCCTCGCGCGGACC 19449
QY 775 CTCACTTACCGGCTCGCGCGGACGCGGCTGCTCCGTCGCGGCTCCGACGGCACAGACC 834
DB 19450 GCCAAGCGCGAGTGCACGCGGAAACGTCACGTTGAGAG---CCGACGCGCGCACACG 19506
QY 835 CCGACGCGCGCGCGCTTCTTCACTACCTGAGAGACAGCTGGAACGCGGACGCGGTCGCC 894
DB 19507 ACCGTGAGAGCGGCTTCTGACCTGCGTGAAGAACGACCTGCGCGGGGCTGCG---CAC 19563
QY 895 GTGCGCGCGGAACTGCGCTTTCGATTCACTGCGTCACTGCGTCACTTCTGCGTACGAG 954
DB 19564 GAGGTGCCGGAATCTCTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGAG 19623
QY 955 CTGAAGCGGAGACACACCGGCGGACCGCGGACCGGTCCCGGACCGCGGACCGCGGCTTC 1014
DB 19624 TTGAAGCGCGAGTGCACGCGGACCGCGGACCGGTCAACCGGATCCGACCGCGGCTG 19683
QY 1015 CTCTTGCAGGACCGGCGCATCGCCCTCGACCAACGAGAGGCTGCTACTGCTGAGCC 1074
DB 19684 GTCTTGCAGGACCGGCGCTTCTGACCTGCGTGAAGAACGACCGGACCACTACTGCTG 19743
QY 1075 CTGACCGCGCGGCGGACGACGACGCGCGCGGCTGCTGCGGAGACGCGCGGAGCC 1134
DB 19744 CTGATGAGAGAGACGACCGGACCGGACCGGACCGGCTGCGGCGGCTTCGCGCAC 19803
QY 1135 CTCAACGCGCTGCGGCTCGCGCGCGCGCGCGGCGGACCGGCGGCTTCGCGGATC 1194
DB 19804 CTGAGGCGCATGCGCGGCGGAGAGC-----CGAGCGGTGC 19839
QY 1195 CCGGAGCGGCGGCGGCTTCGAGCCCGCTGCGCGCGCGCGCGGACGACGAGGCTGCT 1254
DB 19840 CCGGAGCGCGCGGCTGTCAGACGAGGTCTCGGTGAGCTGCGGACGAGCGGAGCGCTAC 19899
QY 1255 CTCAAGCGCATGACGAGTGTCTCAAGAGATCCGGAACGAGGAGTGTGATGATCTGCG 1314
DB 19900 CTGAAGCTGATGACGCTTGCAGAGAGATAGCGCGGGGAGACCTACGAGGTCTGCG 19959
QY 1315 CTGACCAATGCTGACCGCGCGGACCGAGGAGAGGAGGAGGAGGAGGAGGAGGAGG 1374
DB 19960 CTGACCAATGCTGACCGAGCGGACCGGACCTGACCGGAGGAGGAGGAGGAGGAG 20019
QY 1375 CCGGCGCATGAGCGCGGCTTCGATCGGCGCGGCTGCTGAGTTCGCGGAACTGTCGAGG 1434
DB 20020 CCGCGGCTGAGCGCGCGCGCGGCTTCGCGGCTTCTGAGCTTGGGCTCCGAGCGGCTG 20079
QY 1435 AGCGCTGCGCGGAGGCTTCTCAAGATGAGCGCGGACGAGCGGCTGAGTTCAGAGCC 1494
DB 20080 AGGAGCTTCGGAAGGCTTCTGCGCATGACCGGACGAGGCGGAGTGAAGTCAAGCGG 20139
QY 1495 ATCAAGGAGGACCGCGCGCGGCGGAGACCGGAGAGAGAGAGAGAGAGAGAGAGAG 1554
DB 20140 ATCAAGGAGGACCGGCGGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 20199
QY 1555 CTGCGCGGCGGAGAGAGAGGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1614

```

gene
CDS
/transl_table=1
/product="aurd protein"
/protein_id="CAE02599.1"
/db_xref="GI:41322766"
/translation="MVNAVYVLTSPDLSFNITLGPVAVSTSGEPLPIGARORATLSNLT
LARGVVAEAEKLTETALINPAGPPGNAVASLHSYSHLRQLEPPDCPPMGRRIRITRPEQ
GTLVLEVERVDARFEGRLREGRQQLREBGVYPASVLIACFPMRSGVYAEIGDYE
PAVCAALAEELRLATLEALWEALMGHDYVIAIELAVSRFPTRRTGLMVLWA
LCEVROQAEAAVAVNRIKALACQELGVDPGRELSRLFARILRDESVGRACIQS"
complement(3029 . 4546)
/gene="aurb"
complement(3029 . 4546)
/gene="aurf"
/codon_start=1
/transl_table=1
/product="putative acyl-coA ligase"
/protein_id="CAE02600.1"
/db_xref="GI:41322767"
/translation="MLVPLTTADFLDREAETAAARLTADPEHPQAAALFTSYGTELT
RRVRMQGLDGLIDIGDERVAAVVSPNSARLLELYLVATASGRICVAVNRILTPREIS
YLIRCDASVFPVDVDEAVLGSVDVFKRFLVGEOTETVYMRDVPVAPMSRDNNAV
ATLNTSGTSARPKVALTTHRSILWLNMTGYNARIWEGDVYHTLTPTHCNMGVAF
LLAGCAQOVIIRRIDCAEILRRQEGVTLAGAPAVWAVVIAQAARDMFGELPGGR
VRVACGAPVSAQLVSRILENLGWEFLQVGLTETTLITNRLPDAPGGGILTRGA
PALGVRATGSEGVETARSNMVLDVYRDSQANALAEDEMFRGVDGDEGHIAL
FDRVADVITVTCGSESVSEVEDCLOGHPSIAEVAIVGPBERGEEVTKAAVVPADAR
VTEBEVITCYCQOCLARKAFTSVDFRALPRTSNGIKQGRLEPFWAAGDRHITH"
4858 . 5868
/gene="aurf"
4858 . 5868
/gene="aurf"
/codon_start=1
/transl_table=1
/product="putative N-oxidase"
/protein_id="CAE02601.1"
/db_xref="GI:41322768"
/translation="MREBQPIALTTWARGMVEEGISATLIGRLVRAWERAAVYVK
ADILDEMDYDITLVPDPLRIVPFAEHPFLAAEPHORQVLTGMIGYNERVIAEQ
LIAEAPFLVNHGVPFGSDPDLIRKSVQOALVDESFTYHMHAIIDRTRELKISRP
POPELVYTRRLRVLIADMPQEMGMDIVILVWGAIVETCNALLAILARDTIOPMSLT
ITTLHLEDTKHGISVVEVYRELYARNEQORALVRCPLFIAEAEQDLSLIMLET
NNAAGRGAEEIVGDLRSTAGTRLVRPFSGRARKVNEQLGLDAVDFPFRPWSPT
PR"
5990 . 11710
/gene="aura"
5990 . 11710
/gene="aura"
/codon_start=1
/transl_table=1
/product="polyketide synthase type I"
/protein_id="CAE02602.1"
/db_xref="GI:41322769"
/translation="WBLDEHGDSRLERLSAREADRHRLILDTITTSATVATDGRAR
ALDGRFAEGEVGKSGAGKRLRELSAREADRHRLILDTITTSATVATDGRAR
DYDAPAAAIAAADADEPIAIVMGCGRLPGVNSPGEMLVRETTDAISFFPDGRMDV
ETADPDLHGGTITVTRHAGFLYEADFPADFPFALSPEGAVTIDQHRLLIETWEATEG
ARLDPYTSRSGRTGFGVGLMWSEYGARIRNVPPASGAVPVYSGSPVSAAGRLATGEP
EGPARYDTACSSSLVAMHILAAESLRGECSLVPAAGCATYMTSTPKGTITERSRGLAP
DGRASVPSADAGSSWSSEGVGVVLERLSQARNRGRVLAVLNGSAVNDGASNGTLA
PMGQAOQVLRQALVHGLSTADVDVDAAGAGTRLDPTLQALDLYATQGGGRPADOP
LMLSLKSNITGHTQAAAGVGVIKMVAINGVLPKSLHTTETPTVMDSSGAEVLLT
EARMPEPGEGRRAQVSSPGVGNNAVLILEQAPBEPDPAEDGADARPLPWLVS
RGAGLAQAGRLRIPLIEAHPVEDADVAISLAATRAILPDRAVVVVAADRAVAVLA
ALADBOGEOVALVAVNARDRLAFTVFGQSQSPGMAELMSTSPVRSITECEKSL
APVYDWSTLYTLRSGASPTLERDVVQPLFAVMVSLAAMSFCVEBPAAVNGHGS
EIAAAHVASLTALRSDDARIIVASRKALVLSGGGWASVA PVDRLRELERNDGVS
VAANGPRAVYISGPTGDEALIEAFADDLRAKLI PVDASASQVEELRETLRETL
AGIAPRSEVAVSYTGTGPEVDITLIDTEWTLNLETOVQDFTAKRIIDGHTGVE
MSHPRLVTCQERTIDVVAADCVLTPTSLRNEGTRARLITLAGEAHNGITPYDAQV
FDHAGPPEVDITLTAFORRRYWLEBAAPADLASAGDGVDPFLGAAVSPDGSTL
FTGLSARFHWMLIDHGVADVLVVPGTALVDIALAGGQGCVELEELVHAPLILE
QDTRVVDIATGAPDESGRSVTVHARGSDPDPTWTLAEGTLADGRSEDLPGAM
PPAAEATVOLDLCKROLAKNGLDHYGAAFOGLKRGMRIDGVFPVEVLPSGSGAGGY

Db 12462 TCCGCGAGCCGCGCTCTTCTGGCTGACAGACGCGCGCGGAGATCGGCGCGCCG 12521
QY 733 TTCTCTCTTCTCGGCGAGACGCGCGCCCGCTCGCCAGTACTACCTACCTACCGCGTCCG 792
Db 12522 TTCTCTCTTCTCGGCGAGACGCTCCGCGCCCTCGCGAGACGCTGACCTCCGAGCCG 12581
QY 793 GAGGCGCTGCTCTTCTCGGCGAGCTCGGAGGACGACGACCGGAGCGCGCGCCCTTC 852
Db 12582 GAGAACACGCGTCCGCGCGAGCGCGCGGAGACGCGCGCGCGCGCGCGCGCGCGCTTC 12641
QY 853 TTCACTACTGAGAGAGACGCTGAAAGCGCGAGCGGCTCCCGCGCGCGCGCGCGCGCG 912
Db 12642 GAGTGTCTCAGACAGCGCGCTCCGCGACCGCG-----GCTCAGACGCGCGAGCTGCC 12695
QY 913 TTGAGTTCAACTCTCGGCTACGTCGCGCTACCTCGCTACGAGTGAAGCGAGACCGAC 972
Db 12696 TTGAGCTCAACGCGCGCTACGTCGCGCTACTTTCGCGCTACGAGCTCAAGGCGAGTGGCG 12755
QY 973 GCGGACCGCGCGACGCGTCCCGCGACCGCGCGCGCGCTTCTTTCGCGACCGCGCC 1032
Db 12756 AGCCCGAACGTCACGCGCTCCGAAACCCCGAGCGCGCTTGAAGTTTCGCGACCGCGCTG 12815
QY 1033 ATGCGCTTCAGACCAACGAGAGCGTGTCTACTCGTGGCGCTTCGACCGC-----CGG 1086
Db 12816 ATGCGCGTGAACCACTGAGAGAGAACTTACGTCTGGCGCTTGAAGCGCGACCGCGCG 12875
QY 1087 GGCACGACGACGCGCGCGCGCTGCTGCGGAGAGCGCGCGACCTTACCGCGCTG 1146
Db 12876 GCGAGGAGAGACGCGCGCGCGCGCTGCTGAGAGAGCGCGCGCGCGCGCGCGCGCG 12935
QY 1147 GCGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1206
Db 12936 CCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12983
QY 1207 GCGGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1266
Db 12984 TCCGAACAGGTCCAGCATGCGCGCGCGCGCG-----GCGCACTGAGCGACATC 13034
QY 1267 GACGAGTCTCAAGAGATCCGCAACGCGAGTCTGACGAGTCTGCTGACCAATG 1326
Db 13035 AAGGATGCTCTGACGAGTGCACAGCGGAGGAGTACGAGTCTGCTGACCAACG 13094
QY 1327 GTTACGCGCGCGACCGAGCGCGCGCTGCGCTTACTCTCGCGCTGCGCGCGCATGAC 1386
Db 13095 ATCCGCGTCCGCGAGTGAAGACCGCGCTGAGTACCATCTCTGCTCCGCGCGCTCAAC 13154
QY 1387 CCGGTCCCGTACGCGCGCGCTGCTGAGTTCCCGCACTGCGGAGTGAAGCGCGCTGCGCG 1446
Db 13155 CCGCGCGCTTACGCGCGCGCTTCTGCGCGACGAGGCGGTGAGCTTCTGCTCTCCCG 13214
QY 1447 GAGCGGTTCTCAAGATCGGCGCGCGAGCGCGCTGAGTCAAGCCCATCAAGGAGCG 1506
Db 13215 GAGCGGTTCTGCGGAGTCAACGCGGAGCGGTGTGAGAGCGCGCGCATCAAGGAGCG 13274
QY 1507 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1566
Db 13275 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13334
QY 1567 GAGAGGACCGCGCGCGCGAGACTGATGATGATGATGATGATGATGATGATGATGATG 1626
Db 13335 GCGAAGACCGCGCGCGCGAGACTGATGATGATGATGATGATGATGATGATGATGATG 13394
QY 1627 GTCTGCGGATCGGCTTCTGCACTGCGCGCGCGCTTCTGAGGTGAGACTTACGCGCG 1686
Db 13395 GTCTGCGGAGGTCTGCTTCTGCACTGCGCGCGCGCTTCTGAGGTGAGACTTACGCGCG 13454
QY 1687 GTGACACGCTGCTGCTGCACTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1746
Db 13455 GTGACACGCTGCTGCACTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13514
QY 1747 TGGTTCGCGCGCGCGCTTCCCGCGCGCGCTCCATGACGCGCGCGCGCGCGCGCGCG 1806
Db 13515 TGGTTCGCGCGCGCTTCCCGCGCGCGCTCCATGACGCGCGCGCGCGCGCGCGCGCG 13574

QY 1807 GAGATCATGACCGCGCTGAGAGAGCGCGCGCGCTTACTCCGCGCGCGCTCGATGG 1866
Db 13575 GCCATCATGACCGCGCTGAGAGAGCGCGCGCGCTTACTCCGCGCGCGCTCGATGG 13634
QY 1867 TTGCGCTTCAAGCGCGCGCGCGCGCGCTTACTTCTGATGATGATGATGATGATGATG 1926
Db 13635 TTGCGGCTTCTCGCGCGCGCGCGCGCTTCAAGTATGATGATGATGATGATGATG 13694
QY 1927 GGCACGCGAGTTTCGCGCTGCGCGCGCGCGCGCGCTTCTGCGCGCGCGCGCGCG 1986
Db 13695 ACGGCGACGCTATCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13754
QY 1987 TTCAACGAGACCGTGTAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2032
Db 13755 TTGAGGAGTGTCTTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13800

RESULT 12
AP00618_17/c
WPCOMMENT
Sequence split into 61 fragments LOCUS AP00618 Accession AP00618

Fragment Name	Begin	End
AP00618_00	1	110000
AP00618_01	10001	210000
AP00618_02	20001	310000
AP00618_03	30001	410000
AP00618_04	40001	510000
AP00618_05	50001	610000
AP00618_06	60001	710000
AP00618_07	70001	810000
AP00618_08	80001	910000
AP00618_09	90001	1010000
AP00618_10	100001	1110000
AP00618_11	110001	1210000
AP00618_12	120001	1310000
AP00618_13	130001	1410000
AP00618_14	140001	1510000
AP00618_15	150001	1610000
AP00618_16	160001	1710000
AP00618_17	170001	1810000
AP00618_18	180001	1910000
AP00618_19	190001	2010000
AP00618_20	200001	2110000
AP00618_21	210001	2210000
AP00618_22	220001	2310000
AP00618_23	230001	2410000
AP00618_24	240001	2510000
AP00618_25	250001	2610000
AP00618_26	260001	2710000
AP00618_27	270001	2810000
AP00618_28	280001	2910000
AP00618_29	290001	3010000
AP00618_30	300001	3110000
AP00618_31	310001	3210000
AP00618_32	320001	3310000
AP00618_33	330001	3410000
AP00618_34	340001	3510000
AP00618_35	350001	3610000
AP00618_36	360001	3710000
AP00618_37	370001	3810000
AP00618_38	380001	3910000
AP00618_39	390001	4010000
AP00618_40	400001	4110000
AP00618_41	410001	4210000
AP00618_42	420001	4310000
AP00618_43	430001	4410000
AP00618_44	440001	4510000
AP00618_45	450001	4610000
AP00618_46	460001	4710000
AP00618_47	470001	4810000
AP00618_48	480001	4910000
AP00618_49	490001	5010000

AP00618_50 5000001 5110000
 AP00618_51 5100001 5210000
 AP00618_52 5200001 5310000
 AP00618_53 5300001 5410000
 AP00618_54 5400001 5510000
 AP00618_55 5500001 5610000
 AP00618_56 5600001 5710000
 AP00618_57 5700001 5810000
 AP00618_58 5800001 5910000
 AP00618_59 5900001 6010000
 AP00618_60 6000001 6021225
 Continuation (18 of 61) of AP00618 from base 1700001 (AP00618 Nocardia farcinica IFM 1)

Query Match 32.2%; Score 662.8; DB 1; Length 110000;
 Best Local Similarity 60.7%; Pred. No. 4,5e-46;
 Matches 1269; Conservative 0; Mismatches 742; Indels 81; Gaps 8;

QY 1 ATGCGCAGCTTCTGATGACAACTACGACTCGTTACCCACAACCTGTTCCAGTACATC 60
 DB 54159 ATGCGCAGCTGTTGATGACAACTACGACTCGTTACCTAACCTGTACAGCTGATC 54100
 QY 61 GGGAGGCCACCGGGGCAACCCCGCTGCTGTCGCCAACGACG---CCGACTGATCGCG 117
 DB 54099 AGTGAAGGTGAACGGGGTGAACCGAACCGTGTGCGCAACGACGACCGACCGCGCG 54040
 QY 118 CTGCGCGTCGAGGACTTCGACGCGATGCTGTGTCGCCGGGCGCCGCGACCGCGACCGG 177
 DB 54039 CTGAGACTCGCCGCTTACGACAACTATGCTGTCTCCCGCGCCGCGCGCGCGACCGG 53980
 QY 178 GAAAGGAGCTTCGAAATCAGCCCGCGGCGCATCACGACGCGGCTTCCGCTTCGCG 237
 DB 53979 ACCCGGACGTCGGGGCTGTGCGGAGCGGTATCGACGCGGGGTCGACCGCTTGCTCGG 53920
 QY 238 GTCTGCTCGGCCCAACCGGGCATGCGGACTCTTCGCGGAAACCTGTGCGCTCGCCCG 297
 DB 53919 GTGTGCTCGGGGCAACCGGGCATGCTGTGCGCGCGCGGCGGCGGCGTGAACCGCGCAC 53860
 QY 298 GAAACCATGACGCGCGGGGTCTCCGAGGTGCGGACACCGCGGAGGAGCGCTTCGCGGG 357
 DB 53859 GTACCCCGGACGCGCTTCTTCGACCGGGTCGCGCACGACGCGGCGCGGCGCTGTTCGCGG 53800
 QY 358 CTCCCGCTCGCTTCAACCGCGGTGCGCTTACCACTCCCTGCGCGCACGCA---CTTCCC 414
 DB 53799 CTGCGGAGGATTTCAACGCTGTGCGGTACCACTCGCTGTGCGCGCTGAGCGCGCTGCC 53740
 QY 415 GACGAGCTCGAACCCCTCGCTGAGGCGACGCGGGGTCTCATGCGCTTCGCGCACCGC 474
 DB 53739 CCGCGCTCGAGGTGACCTGCCCCGACGCGCGGCGGCTGATCATGAGGGGTGCGGCATCG 53680
 QY 475 GAGAAAGCGGTGAGGGGCTGCAAGTTCCACCCGGAATCCATCGGCAAGCACTTCGCGCG 534
 DB 53679 CACCGCGCGGAGTGGGTGTGCAAGTTCCATCCGAGTGGTGGCGCGGAGTTCCGCGGCG 53620
 QY 535 GAGATCATGCGCAACTCCGCGACCTCGCCCTCGCCAC-----CACCGGAGCAAGCGGC 589
 DB 53619 GCGCTGTGCGCACTTCGCGGAACCTGCTGCGGACGACGCGCAAAACCCGCGCGGACG 53560
 QY 590 ACCGAGCGCACTCCCGCTACGAATCCACGTGCGCGCGCTGACGCTGCTCGGACGCGC 649
 DB 53559 CGTCCGCGACACCCGCGACGCGCGCGCTCGCGCGCGACGACGCGCGCGGTGCGCG 53500
 QY 650 AAGAGTACGCGCGCGCTGCTGCGCG-----677
 DB 53499 CTGCGGACACGAGGTGATGAGCGCGCGGTGACACCGAGGCGCGGTTCTGCGGCTTAC 53440
 QY 678 -----CGAGGAGCAACGTTCTGGCTGAGACAGAGCTCCGCTCGAAGGCGCGCTCGCG 732
 DB 53439 GCGCGCTGCGCGACGCGGTTCTGCTGACACGCAACAGTCAAGCGCGGGGCTGACCGG 53380
 QY 733 TTCTCTTCTTCTCGCGACGACCGCGGCGCTGCGCGAGTACCTACCTACCGGCTGCC 792
 DB 53379 TTCTCTTCTTCTCGGTTGACGCGAGCGGACCGCTGCGGAGGTGAGTTCGCTTACCGGCTCG 53320

QY 793 GACGGGTGCTCTCCGTCCGCGCTCGACGAGCAACGACCGGACCGCCCTTC 852
 DB 53319 GCGGACGTGTGCGGGTTCAGGACGCGCGGAAACCGGACGCTCCGCGGAGCGTGTG 53260
 QY 853 TTCAACTTACCTGAGAGGACAGCTTCGAACGCGGAGTCCCGCTCGCCCTGAACTGCC 912
 DB 53259 GGTATCTGTCCACCCAGCTGCGCA-----CGCGCAGCTTCAGGTGCGCCGACCTGCC 53206
 QY 913 TTGAGTTCAACCTCGGCTACGTGGCTACCTGCGCTACGAGCTGAAGGCGGAGACAC 972
 DB 53205 TTGACTTGTGTGCGGGAATGCTGGCTTACCTGTGCTTCAAGATCAAGGCTGCGG 53146
 QY 973 GCGACCCCGCGACCGGCTCCCGCACCCCGACCGCGCGCTTCTTTCGCGGACCGCGCC 1032
 DB 53145 GCGCGGGCGCGGACCGGCGCGGCGCGCACACCGGAGCGCGCATCTTCGCGACCGGCTG 53086
 QY 1033 ATGCGCTTCGACCAACGAGAGGCTGTCTTACTCTGTGCGCTTCGACCGCGGAGCAC 1092
 DB 53085 ATGCTGTGACCATGTGCGGGGACAGACCATCTGTGCGCTTAC-----GACGAC 53032
 QY 1093 GACGAGCGCGCGCGCTGCTGCGGAGAGACGCGCGACACCTCACCGCGCTGCGCGTC 1152
 DB 53031 GACACCGCGCGCATGATGATGCTGCGGAG-----ACCGGCGGAGTGTG 52987
 QY 1153 CGCGCCCGCGCGAGCGGACCGCGCATGTCTTCGAGATCCCGAGCGCGGCGCGCG 1212
 DB 52986 GAAACGCTGCGCGCTGCGGCGGATTCGCGGCAATTCGCGATGATGATTCGCGCGCTCG 52927
 QY 1213 TTGCGCCCTTCGCGCGCGCGCGCGCACGACGAGAGCGCTTACTCAAGCGCATGACG 1272
 DB 52926 GTGCGCGCTGTGTAACGCGGCGCGCGACCG-----TACTCCCGGACATGCGCGCG 52873
 QY 1273 TGCTTCAAGAGATTCGCGACGCGGAGTGTGATGAGATCTGCTTCAAGCAATGTGTACC 1332
 DB 52872 TGCTTCAAGATTCGCGCGCGGCGGAGTGTGATGAGATCTGCTTCAAGCAACCTTACC 52813
 QY 1333 GCGCGACGAGGCGGACGCGCGCTGCGCTTACTTCGCGCTGCGCGCGCATCAACCGCGTC 1392
 DB 52812 GTGAGCGCGGATTCGCGGCGGCTGAGACTTCTATTCGACGCTGCGCGCTGCAATCCGCG 52753
 QY 1393 CCGTACGCGCGCTGTCTGAGTTCCCGAATGTGCGGTGAGCGCTTCGCGCGGACGCG 1452
 DB 52752 CCTATGCGCGCTTACCTGCGCTTTCGACGACCTGAGATGCGCTGCTTACCGGAGCGG 52693
 QY 1453 TTCTTCAAGATTCGCGCGCGGCGGCGGCTGTGAGTCCAGGCTTCAAGGAGACCGCGCC 1512
 DB 52692 TTCTTCAAGATTCGCGCGGCGGCGGCGGCTGTGAGGAGCAACCTGATCAAGGCGAC 52633
 QY 1513 CCGGCGCGGACCGCGGAGGAGCGAGCGGCTCGCGCGGACCTGACCGCGCGGAGAG 1572
 DB 52632 CCGGCGGCGACCGCGCGCGGAGCGAGCGGCTGCGCGGAGTTGGCCGACGATCCGAG 52573
 QY 1573 GACCGGCGGAGCACTGATGATGCTGACGCTGTGCTCGGACGACCTTCAACAGCGTTCG 1632
 DB 52572 ACCCGCGCGGAGCACTTATGATGCTGACGCTGTGCTGCGACGACCTTCGCGCGGCTGCG 52513
 QY 1633 GCGATGCTGCTCGTTCACGCTGCGCGCTTTCGAGTGAAGCACTTACGCGCGCTGAC 1692
 DB 52512 CAGATCGGACGCGTGAACGATACCGAGCTGATGAGCGAGCGAGTTCGACGCTGAC 52453
 QY 1693 CAGTGTGTGACCACTCCGCGGACGCGCTGCGCGCGCGGACGACGACCGCGCGCTGCG 1752
 DB 52452 CAGTGTGTGACGCGTGTGCGGACCTGCGCGCGGAGTTCGACGCTGACGCTGCGT 52393
 QY 1753 CGCGCGCGCTTCCCGCGGCTTCATGACGCGGCGCGGCAAGAGCGGACGATGAGATC 1812
 DB 52392 CGCGCGCTGCTTCCCGCGGCTTCATGACGCGGCGCGGCGGAGTGTGCGAGATGAGATC 52333
 QY 1813 ATGACCGCGCTGAGGAGAGCGCGCGGCGCTTACTTCGCGGCGCTGAGATGTTGCGC 1872
 DB 52332 ATGACGAGCTGAGGAGCGCGCGCGGCGGCTTACTTCGCGGACGATTCGCGCTTCGGA 52273
 QY 1873 CTCAGGCGCGCGCGGACCTGACGACATGCTATCCGACCAATGCTGTGCGGCGGACGCG 1932

Db 52272 CTGGGGCGGCAACCGCGCACTCAATCGTATCGACCGCGGTGGCGCAAGACGGCGC 52213

Qy 1933 GCCGAGTTTCGGCGTCGCGGGCGGCGATCGTCCCTCCGACCAAGAGAGATTACAC 1992

Db 52212 TGCGGATTCGGGGCGGGCGCGCGATGCTGTCGATTCGATCCGCGGAGAGATATCGC 52153

Qy 1993 GAGACCGTGTGTAAGGCGCGCCGATGATGTCACCGCCCTCGACGCGACGCGC 2044

Db 52152 GAGATGTTGTGAAGGCGCGCGACCCAGCGCGCGCCACCGCGCGCGC 52101

RESULT 13
AP005025/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AP005025 298300 bp DNA linear BCT 10-MAY-2003
Streptomyces avermitilis genomic DNA, complete genome, section 5/30.
AP005025 BA000030
AP005025.1 GI:29604584
Streptomyces avermitilis MA-4680
Streptomyces avermitilis MA-4680
Bacterial Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

1 Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kikuchi,H., Shiba,T., Sakaki,Y. and Hattori,M.
Genome sequence of an industrial microorganism Streptomyces avermitilis: deducing the ability of producing secondary metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001) 21477403 11572948

2 Ikeda,H., Ishikawa,J., Hanamoto,A., Shinose,M., Kikuchi,H., Shiba,T., Sakaki,Y., Hattori,M. and Omura,S.
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003) 22608306 12692562

3 (bases 1 to 298300)
Omura,S., Ikeda,H., Ishikawa,J., Hanamoto,A., Takahashi,C., Shinose,M., Takahashi,Y., Horikawa,H., Nakazawa,H., Osonoe,T., Kushida,N., Director-General of Biotechnology Center, Shiba,T., Sakaki,Y. and Hattori,M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center/ 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan (E-mail:bic@nte.go.jp, URL:http://www.bio.nte.go.jp/, Tel.81-3-3481-1933, Fax:81-3-3481-8424)

This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Osonoe(*4), Norihiro Kushida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.

*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.kitasato-u.ac.jp.
Location/Qualifiers
1. .298300
/organism="Streptomyces avermitilis MA-4680"

/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227892"
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermitilinus"
gene
208. .921
/gene="orf-1"
/gene="orf-1"
/gene="orf-1"
/note="SAV944
not involved in biosynthesis of avermectin"
/codon_start=1
/transl_table=11
/product="reductase"
/protein_id="BAC68654.1"
/db_xref="GI:29604585"
/translacion="MCVDRDGRGAGVTDLADTARARAMEPVADVSDGAMERPAER
VAETGVVDLAVNMAGIGMAGRFDTSDVDMQRTGVMIMGICRLIGRMARQ
GGHIVTASAAAFQPTPAVPAVATSKAVALMSECLRAEFAGVGVSVCGPVRS
PASAHFAGVPRLEOBRALPARGCSAEKVAALVLSVARDAVVTATBARLSRL
MSRFTPRLAAVARMDPPS"
979. .2217
/gene="aveB1"
/gene="aveB1"
979. .2217
/note="SAV945
avermectin gene cluster"
/codon_start=1
/transl_table=11
/product="glycosyl transferase"
/protein_id="BAC68655.1"
/db_xref="GI:29604586"
/translacion="MSDHLFMSAPFMGHPFVPSLVAABELVHRGHVTVTGAEADA
VRSQADPRRYSABGVDMYRLMTEABENATPMTLYDEGSMLSVEHVCKDVPDL
VADTASLNVSRVLAASWSRPAMTVIPLFASNGRFTSQSVLDPDSAOVSAAPRFS
EOMELFGALAPRLAEELVSRGITEPVDDFISGDEPNVCLPRAFOYAGTFPERF
AVGSCGLGRKRGTEIPGSGHPVLISLGVFNROLSFFRTFVRAFTVDVHVVIS
LCKGVDPVLRPLPNNVEVHRVPRHVALEHARALVTGCTGVSWEALHAGCPVLMV
LSRDQVGRRIAEIIGRMVQPEEVTTLLRRHVLDIISDAITRQVRQMRATVEA
GGALRAADBERFERLRTTRH"
complement (2123. .3190)
/gene="aveB11"
complement (2123. .3190)
/gene="aveB11"
/note="SAV946
avermectin gene cluster"
/codon_start=1
/transl_table=11
/product="dtdp-glucose 4,6-dehydratase"
/protein_id="BAC68656.1"
/db_xref="GI:29604587"
/translacion="WTGGAGPFGSHFRRRLTGAYPATGAEVVVLDKLTVAQLENT
AVLGSPLSILFVHGDICGPELVADIMDSDVVHFAASHVDRSVAADAERFRTVLG
THTLLRAATVADVRRFVISTDEVGSLDSGSMTDAPLEPNSPVASAKASDLARS
FRTTGLPYLITRCSNNYGHQFPKLLPRFTHLNLTGYVLYDGEVNRDMLHVD
HORGLAIVASERGRPEGIYHIGGTLNSRELTRALLDLYGWSWVEPVTDRKGDRR
VSLDISKSAELRGVAPVHPEBGGLAQTYQVMYENTLMBPLTARELPEVSDGASAE
ASRPUPAGRRPRPWPASAS"
complement (3187. .4086)
/gene="aveB111"
complement (3187. .4086)
/gene="aveB111"
/note="SAV947
avermectin gene cluster"
/codon_start=1
/transl_table=11
/product="glucose-1-phosphate thymidyltransferase"
/protein_id="BAC68657.1"
/db_xref="GI:29604588"
/translacion="WKGIVLAGCTGSRLYPLTRALSQKLLPYDCKPMIYPLSLMLG
GIKDLIVSSPDHLQFRRLPDGSRILANTDYAAQQRPGGIAEFLIGADPTGDDV
SLVLDNITFHGWSFHLRSHRTVDGCVLRCGYAVTDEBRYGVEVDASGKLLSVEEK

Db	266894	ACTGCGGTTCCGTCAACCGTCAATCGGAGCCGAACTCCGACGCTGCTGCTGTTCGCG	266835
Qy	1025	ACCGGCGCATCGCCCTCGACCAACAAGAAAGCTGTCTACTCGTGCCTCGACCGCC	1084
Db	266834	ACCGGCTGATCCGAGTGGACATCAAGACGGA-----GCCACTTAGCCG	266790
Qy	1085	GGGCGCCAGCACGAGCGCGCCCGGCTGCGTGGAGACGCGCGAACCTTCACCGGCC	1144
Db	266789	TCTGTCTCCCGAGAACACCCCGCAGGGGCAACGAAAGCGGCGCATCGACTTCAGCGCG	266730
Qy	1145	TGGCGGTGCGGCGCCCGCGAGCGCACCCCGGCATGTGCTTCGGGATTCGCCAGGCGG	1204
Db	266729	CGATGCCCCAGCTCAAGTTTCGTGTCTTCGAGCCAAAGCCGAGCCCGCGGCGACCG	266707
Qy	1205	CGAGCGGCTTCGAGCCCGCTG---CCGCGCGCGCACACAAGGACGCGCTACTCAAG	1261
Db	266699	CACCCCGCTTCATGCGGTGAGACCGTGAGCTGTGTCGCGACCGTACGACTTACTGCGG	266610
Qy	1262	GCATCGACGAGTGCCTCAAGAGAGATCCGACCGGAGTGTGACGAGATCTGCTTACCA	1321
Db	266609	ACATCGGAGACTGCAAGCGGGAACCTCAAGGACGGACACAGTTTACGAGATCTGTACCA	266550
Qy	1322	ACATGCTACCGGCGCGACCGAGGCGACGCGCTGCGGCTCTACTCGCGCTGCGGCCA	1381
Db	266549	ACGCGGCGCAGGCTGCGCGCCCGTACGACCCCTACCACTTCTTACCGGATACGCGCTC	266490
Qy	1382	TCAGCCCGCTCCGATCGGCGCCCTGCTCGAGTTCGCCGAACTGTGAGTGTGAGCGCCT	1441
Db	266489	TGACCCCGGCGCGTACGCGGCTTATCTGCGTTGCTGTACCTTGATGTGACCGGCTCGT	266430
Qy	1442	CGCCGAGCGGTTCTCTACGATCGGCGCGACCGGCGCGTCAAGTCCAAAGCCCATCAAG	1501
Db	266429	CCCCGAGGTTTCTTGCGGATACCCGCGAGCGGCTCGCGAGGCGACAGCCCGTTCAAG	266370
Qy	1502	GGAACCGCCCCCGGGGCGGACCGCGGAGAGAGACGAGGCTCGCGCGCGACTGCGCG	1561
Db	266369	GACCGCGCGCCCGCGCGCGAGCGCGCGAGAGAGCCCGGCTTACGAGACGCGCTCACCA	266310
Qy	1562	GCCGGAGAAAGACCGGGCGGAGAACCTATGATGTGTGACCTGCGCTGCCAATGACCTCA	1621
Db	266309	CGGACGACAAAGACCCCGGAGAGAACTGGTATGCTGTGACCTGCTTCGAAAGACTTGG	266250
Qy	1622	AAGCGTCTGCGGATCGGCTCGCTCGTCCAGGTGCCCGGCTCTTCGAGTGGAGACTTACG	1681
Db	266249	GCGGGTCTGCGGAGCGGGAACGTTAAAGTCAACCGCTGATGCGGACCGAGACGTTGGG	266190
Qy	1682	CGCCGCTGACCACTGTGTGTGACCATCCGGGGAACGCTGCGGCCGCGACCAAGCACCG	1741
Db	266189	CCACCGTGCACCAAGTTGTCTCCACCGTGAAGGGGCGGCTGCGCGAGGCGATGAGAGCGG	266130
Qy	1742	CGGCGGTGTAAGCGCGCGCTTCCCGGCGGCTCCATGTAACCGCGCGGCCAAGAAACCGCA	1801
Db	266129	TGAGCTGCGTACGTGTCTGTCTTCCCGGAGTTTGGTGAACCGGAGCGCGAAACTGCGCA	266070
Qy	1802	CCATGAGATCATTCGACCGCTGAGAGAAAGCCCGCGGAGCGTCTACTCGGAGGCGCTCG	1861
Db	266069	CGATGGAAGTCAATCGACTCGCTGGAGACCGAGGCGCGCGGGGTACTCTCGAGACCATCG	266010
Qy	1862	GATGTTGCGCTCAAGCGGCGCGCGACTTACATGCTCATCCGACCATGTTGCTGG	1921
Db	266009	GCTACTTTCGGGTGACATGGGGGGCGGACCTCGCCATCGCAGCGCACGCGCGTGTCA	265950
Qy	1922	CCGACGGCCAGGCGGAGTTTGGCGTGTGGCGGGCGCATCTGTTCCTCTCGACCAAGAG	1981
Db	265949	CCGACGGGAGATGCACTTCGCGCGCGGCGGTGCGATGTGCTCTCGACTCGATCCGATCG	265890
Qy	1982	AGAGTTTCAACGAGACCGTGTAAAGGCGCGGCCATGTGACACCGGCTCGACGCGAGCG	2041
Db	265889	GCGAGTACGACGAGATGCTGTGTGAAGACGCGCGCACCGATGCGCGCCACCGCGACCGGA	265830
Qy	2042	CG 2044	

Db	265829	TCG	265827	
RESULT 14				
LOCUS	AP005217			
DEFINITION	AP005217	300750 bp	DNA	linear BCT 24-JUL-2003
ACCESSION	Corynebacterium efficiens YS-314	DNA, complete genome, section		
VERSION	4/11.			
KEYWORDS	AP005217 BA000035			
SOURCE	AP005217.1	GI:23492722		
ORGANISM	Corynebacterium efficiens YS-314			
REFERENCE	Corynebacterium efficiens YS-314			
AUTHORS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Corynebacteriaceae; Corynebacteriaceae; Corynebacterium.			
	1 Nishio, Y., Nakamura, Y., Kawarayashi, Y., Usuda, Y., Kimura, E., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K. and Gojobori, T.			
	Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens			
	Genome Res. 13 (7), 1572-1579 (2003)			
JOURNAL	2 (bases 1 to 300750)			
MEDLINE	12840036			
PUBMED	12840036			
REFERENCE	Kawarayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and			
AUTHORS	Director-General of Biotechnology Center.			
TITLE	Direct Submission			
JOURNAL	Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-Ku, Tokyo 151-0066, Japan (E-mail: bio@nite.go.jp, Tel: 81-3-3461-1933, Fax: 81-3-3461-6424) Kawarayashi, Y is officially affiliated with the National			
COMMENT				

FEATURES

Institute of Advanced Industrial Science and Technology, Tsukuba
Ibaraki, 305-8565 Japan
Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the
National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan
Itoh, T. is at the Japan Biological Information Research Center,
Koto-ku, Tokyo, 135-0064 Japan
Yanagishi, A. is at Tokyo University of Pharmacy and Life Science
Hachioji, Tokyo, 192-0392 Japan
Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co.,
Inc., Kawasaki, Kanagawa, 210-8681 Japan
The other authors are at the National Institute of Technology and
Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.
Location/Qualifiers

```

CDS
    /organism="Corynebacterium efficiens YS-314"
    /mol_type="genomic DNA"
    /strain="YS-314"
    /db_xref="taxon:196164"
    252_620
    /note="CB0885"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"
    /protein_id="BAC17695.1"
    /db_xref="GI:23492723"
    /translation="MDQAPSGCLVDCVHGFI FDTVPYMLLPVGISGNRKIVIDLFN
HYFLINKPRILPLPMSGCCGLFGWVIGLFTPMTHYVVFCGEFDDPSGRTLAQC
LTHASRTMKPCLQGGLVANG"
    529_1936
    /gene="rrna_16S"
    529_1936
    /gene="rrna_16S"
    /gene="rrna_16S"
    /product="16S ribosomal RNA"
    /note="CE-RNA01"
    complement(2042..2215)
    /note="CB0886"
    /codon_start=1
    /transl_table=11
    /product="hypothetical protein"

```

[illegible]

		EVARLQRMILDALSTAPSLTKEGLACGRMTRLRELQVDEQVAVRSLGVSSGL RGVPDPLADDDGDYIAFPVADEMULATAEQLAHIRGMMTOTIYAMLEADE KNKPHLLSPASRRDLPEETRAMILOSLPVRVEEDLEALFHHPLAASRIPEETLRH LLEEARMKAVAGGCVTSPPARVLTPAGRVIEPIDIAVAVAAPKPVDFIYOADVTY WYPGELOEMOKIIGOLDIESPGLSAYVRIEASLRHMMDGLTTEIEDFLTOHSA TGLPOSVAILODIARRHGTHSGPALCYLRSDDPALHSAAEAACALGIROIAPFYA VSNMABLIEVITLRKRAGQPVAEDLAGSDDLAPRPAPPAEQRSACGPDSLRIRA AVALARREAVARTGTVSOPPLVALQAARVGORTVTLGFVDKQGVAVHRIVKLPTNA GOVDAIDESTGTVHRFMLRITEVID"
CDS		/note="CE0893, similar to AUs893924-91 CAC31110.1 percent identity: 50 in 62 aa" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAC17703.1" /db_xref="gi:23492731" /translation="MGNNANVERKKGYVDPAWDHPDBSEGHVVTELIPAYGASSPMCD DMEFPVATTGTYHPIRYRINR" complement(12229..12822) /note="CE0894, similar to AX133781-1 CAC38957.1 percent identity: 71 in 195 aa" /codon_start=1 /transl_table=11 /product="conserved hypothetical protein" /protein_id="BAC17704.1" /db_xref="gi:23492732" /translation="MLFMGRHSAKTATPFVKMAATTVAFGAATAVAFAPASAPSDM DRINCESGGNMAITNGYHGLOFSDPTWAYGDEBPAPYAOTREOOLAVART LAGCKGMAMPACSSKLGINSAPTRENLPAPAPAPAPAPATRYATVDINTNPVCCS DLNTVQGIIYDAVSGTLAQIGVOVPAEIOMHTVAFGR"
CDS		13342..13725 /note="CE0895, similar to AE006977-4 AAK45135.1 percent identity: 52 in 127 aa" /codon_start=1 /transl_table=11 /product="putative cold shock protein" /protein_id="BAC17705.1" /db_xref="gi:23492733" /translation="MPGVTKMKVIBERGGFVSNPDGEDCFVKGQVLPKGVTELHOGO RIEPRFARGRKGPOLARKYLETPRRRRQHTYKPRELNIGISDMWTMLEGTQPALLK GHYPRKXVEQVAKILRAVAKELEA" complement(13881..14408) /note="CE0896" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAC17706.1" /db_xref="gi:23492734" /translation="MASRRTKRKNFIQTATLLVAVVVVISVFQNMMNNRPERPEPETSITASPAGIEVFPSYCIEPGECAENEVPLEVGADDELRLTIPEIIDHDVYL LTIVDDPANDEFHTSYETTEVIVPGSDVPTBEGGERRLVVEVTSWMIGDNUGE ETPYTWSLATVEO"
CDS		complement(14424..15254) /note="CE0897, similar to AB001874-1 AAF09704.1 percent identity: 40 in 267 aa" /codon_start=1 /transl_table=11 /product="putative glutamine cyclotransferase"
Query Match	23.2%; Score 478.4; DB 1; Length 300750;	
Best Local Similarity	57.6%; Pred. No. 2.4e-31;	
Matches 1164; Conservative	0; Mismatches 676; Indels 180; Gaps 10;	
OY	2 TGC GCACGCTTGATCGACA CTACGACTGTTACACCACA CCGTTCACGTACATCG 61	
Db	184889 TGCGGTA CTACTCATCGACA ATTATGATTCCTTCA CCTTCAACCTCGCACA GTATGTCG 184948	
OY	62 GCGAGGCCACCGGGAACCCCCCGCGTCGTCGCCAACGAGCGCGACTGTCGGCGCTCG 121	
Db	184949 AGGAGGTACCGGTGCGGCA CGACGCGTGGTCCCAATGATG-----CGCAATCG 184999	
OY	122 CCATCGAGACTTCGACCGCATCTGCTGTCTCCC CGGCGCCCGCGACGCCGACCGCGA AC 181	

Db	185000	ACGAAACGCTGTTCCAGACGCGCGATCATCTCAACCCGGCCCCGGCCACCCCGGGGTGGCG	185059
OY	182	GGGACTTGGGAATCAAGCCGCGGGCGATCAACGACAGCGGCTGGCCGTCTTCGAGCTCT	241
Db	185060	CGGAATTTGGGAGCTGCGCCCGGGGTGATCGAAGCTGTCTGGTGCCTCGGCTCTCGGGGTG	185119
OY	242	GGCTGGGCGACAGGGGCAATGCGCCAGCTCTTGGCGGGAAACCGTGGGCTCGCCCGGAAC	301
Db	185120	GCTTGGGCGATAGGGGCATGCGGCTCGCCACGCGCGCGGGGGCGCGCCGATCC	185179
OY	302	CCATGACGCGCCGGGTCTTCGAGGTGCGGACAGCCGGCAGAGACGTCTTCGGGGGCTCC	361
Db	185180	CGGTGACGGTCAAGGTCAACCCGATCCACACGAGGGGTCCAGACTTTTGAGCATCC	185239
OY	362	CCTGGCCGTTCACCGCGGTGCGCTACACTCTCTGGCCGCGACGACTTCCCGACGAGC	421
Db	185240	CGCCCAAGTTCACGCGCGCTCCGTACACTCTCTGTGGCCACCGAATCGGCCCGGAAC	185299
OY	422	TCGAACCCCTGCGCTGGAGCGAGACGGGGGTGTCATGGGGCTGGGGCACCCGGAAGC	481
Db	185300	TGGAGGTCAACCGCGAGAGACCGGGGACGGGCTGATATCAATGGCGCTGGCCACCGGACTGC	185359
OY	482	CGCTGTGGGGCGCTCCAGTTCCACCCGGAGTCCATCGGACGACTTCGGCGGGAGATCA	541
Db	185360	CCGATGGGGGTGTGCAAGTTCCACCCGGAAATCATCGCGGGGAGTTCCGGGCACCGGATCA	185419
OY	542	TGGCCAACTTCCGGCACTTCGCGCTTCGCCACACCTGGGACGAGGCGCACGGGGCGACT	601
Db	185420	TGGCAATTTCTTGAGCCTTGGC-----GGTTCGACAAAGCGACCGGTGGAGA	185467
OY	602	CCCCGTACGAATCCACGTCGCGCGCGCTGACGTCGTGCGGAGCGCGGAAGATACGCC	661
Db	185468	TCACCGAGCATGTGTGAGAGACAAGGTGACCCGGCGGCGGTTCGAGAGCTCTTGG	185527
OY	662	GCGGCTGCTGCTGCGCGGAGGCGACCACTTCTGTGACAGCAAGACTCCGTCTTGAAG	721
Db	185528	CGGGGTGGAGGACCG-----TTCTGGCTGATG	185557
OY	722	GCGCCTCGCGCTTCTTCTTCCTCGGCGAGACCGGGCCCGCTCGCGAGTACTCACT	781
Db	185558	ATCCGACAGGACCAACTATATATGGGTAAGCTTCGCGGCGCATGACGATTCGACCC	185617
OY	782	ACCGGTGCAGACGAGCGGTGTCCTCGTCGCGGCTCCGACGACACAGACCCGGAGCG	841
Db	185618	ACCGGGTGGGGAGGGGGAGCTGTTCGACTG-----	185648
OY	842	GGCGCCCTTCTTCACTACCTGAGAGACACTCGAACCCGACGGGTCCTCGTGCCTC	901
Db	185649	-----GCTGCGTATATGTGGGTCCCAACCGGTGGCCC	185683
OY	902	CCGAATCGCCTTGAGATTCAACTCGGCTAGTGGGCACTACCTCGGCTACGAGTGAAG	961
Db	185684	CGGGGCTG-----GTTTTCCGCTGGGGTGGGTGGGAATACCTCGGCTATGAGATAAAG	185737
OY	962	CGAGACCAACCGGCGACCCCGCGCACCGGTCTCCCGACCCCGACGCGCGGATTCTTTTG	1021
Db	185738	CGGAATGCGGGGTGAGCAATCGGACAGCTCATTCGATCCCGATGCGCACCTGATTTTG	185797
OY	1022	CCGATCGGCGCATTCGCGCTTCGACACACAGAAAGCTGCTACTCTGTGGCCCTGACC	1081
Db	185798	CCGACCCGGGCAATTCGCATTCGAACCC-----GGCGCGCTGTGGCTCATGAGGCTTCGTG	185851
OY	1082	GCGGGGGCGACGACGAGCGGCGCGCGCTGTGGCTGGGAGACGGCGGAGACCTTCACCG	1141
Db	185852	AGCAGGGGGAG-----TGGTTGCGGAGATGATCGCGGCTCT-----	185888
OY	1142	GCTTGGCGCTCGCGCCCGCGCGAGCCGAAACCCCGCCATATGCTTTCGGATCCCGAAG	1201
Db	185889	---GGGGAGCTGGCGCCACCCCGTGC-----	185913
OY	1202	CGGCGGCGGCTTGGCGCCCTTGGCGCGCGCGCGCATCAGACAGCGCTTACTCAAGC	1261

Db	Accession	Sequence	Length	Source
Db	185914	-GCTGCGCCCGCCCGCCCGCAGCTCACCGTCCGGGATGACCGCGACAGCTACCTGGACA	185971	
Qy	1262	GCATTCAGCAGTGCCTCAAGAGATCCGCACGCGAGTCGTACAGAGATCTGCGTACCA	1321	
Db	185972	TGATTCGCCCGCGCCACGAGAGTTGATCACCCCGCGGAGTCTATGAGATCTCCTBACCA	186031	
Qy	1322	ACATGTCACCGCGCGCGACCGAGGCGAGCGCCCTACCTACTCCGCGCTGCGCGCA	1381	
Db	186032	CCAGCTGCGCGCGAGGTGGAGGTGATTCCTCTCGCGCTATCTGCGCTGCGGCGG	186091	
Qy	1382	TCAGCCCGCTCCCGTACCGCGCGCTCTGAGTTCCCGAACTGTGCGGTGAGCGCT	1441	
Db	186092	CCAAACCCACCTCTGATGATCATTTCTGACAGCTGCGGGGAGATGGCGGTGCTGAGTCTG	186151	
Qy	1442	CGCCCGAGCGGTTCTTCAAGATCGCGCGCGACGCGCGTCAAGTTCAGGCCCATCAAG	1501	
Db	186152	CGCGGAGACGGTTATCAACATCGACGCGCTCCGGACGTGTGAAATCCAAACCATCAAG	186211	
Qy	1502	GGACCCCGCGCCCGGGGCGGACCGCGAGAGAGAGAGAGCGGCTCCGCGCGCGACTGGCGG	1561	
Db	186212	GCACCCCGCGCGCGGGGCGAGTACCGAGCAGAGAGATCGCGCTTGATTCGCCGATCTTCACCG	186271	
Qy	1562	GCCGGAGAGAGACCGCGCGCGAGAACCTGATGATGTGCACTGTGTCCGCAACGACCTCA	1621	
Db	186272	ACAACCCCAAGACCGCGCGCGGAACTCATGATGTGCACTGTGTCCGCAATGATCTCG	186331	
Qy	1622	ACAGCGTTCGCGCGATCGGCTTCGTCACGTCGCGCGCTCTTTCGAGTGTGAGACTTACG	1681	
Db	186332	CCCGCGGGGCGCAACCGGCAACCGGTGACGTGAGAGAGAACCTTTTCACGCTCGAGACTTACG	186391	
Qy	1682	CGCCCGTTCACAGCTGTGTGTGACATTCGGGAGACGCTGCGGCGCGGACACGACCG	1741	
Db	186392	CCACCGTTCACAGCTGTGTGTGACATTCGCGGAGCTGAGAG--GTTAAGAGCCCA	186448	
Qy	1742	CGCGCTGCTGACGCGCGCGCTTCCCGCGGCTCCATGACCGGCGCGCCCAAGAGCGCA	1801	
Db	186449	TTGACTGCGGTTCGGGCGCGCGTTCGCCGCTGTGTTTCATGACCGGTGACCGAAGATCCGCA	186508	
Qy	1802	CCATGAGAGATATGACCGCGCTGAGAGAGAGCGCCCGGGGCGCTACTTCGGGCGCGCTCG	1861	
Db	186509	CCATGAGATATATGACCGAGCTGAGAGACCGCGCGCGCTGTGTGTTATTCGGGTGCGCTCG	186568	
Qy	1862	GATGCTTCGCGCTGACCGCGCGCGCGCGACTCAGCATGTGTCATCCGACCATCTGTCGTG	1921	
Db	186569	GCTATTTCTCTTGTGATGGGCGGTGTGATCTGTGATGTGATTCGCAACCGTGTCTACA	186628	
Qy	1922	CGAAGCGCACGCGAGATTGCGCGTGGCGGGCGGATGTGTCTCTTCGACCAAGAGG	1981	
Db	186629	CCCCCGGCGTCTGTGAGTACCGGGGTGCGCGGGCGCATCTGGCGCTGTGCGACCCGCGG	186688	
Qy	1982	AGGAGTTTCACGAGACCGTGTGTAAGCGCGCGCGCATGT 2021		
Db	186689	CAGAGTGGAGAGATCCGGGTGAATGAGGCGCTCT 186728		
RESULT 15				
LOCUS	A48326	4496 bp	DNA	linear PAT 07-MAR-1997
DEFINITION	Sequence 4 from Patent WO9601901.			
VERSION	A48326			
KEYWORDS	A48326.1 GI:2302119			
SOURCE				
ORGANISM	Streptomyces pristinaespiralis			
REFERENCE	Streptomyces pristinaespiralis			
AUTHORS	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
	Streptomycineae; Streptomycetaceae; Streptomyces.			
	1 (bases 1 to 4496)			
	Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,			
	Barriere, J., Debusche, L., Fanechon, A., Paris, J. and			
	Dutrieu-Rosset, G.			
TITLE	STREPTOMYCES PRISTINAE SPIRALIS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS			
JOURNAL	Patent: WO 9601901-A 4 25-JAN-1996;			

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 16:46:09 : Search time 1245.11 Seconds
(without alignments)
9798.799 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061
Sequence: 1 atgcgcacgcctctgatcga.....ccgtggcgsgcgcccgatga 2061

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412.

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2061	100.0	2061	4 AAF66409	Aaf66409 4-amino-4
2	2061	100.0	2061	8 AAL50180	Aal50180 S venezue
3	1773.8	86.1	3305	8 ABZ69799	Abz69799 Plasmid P
4	1772.2	86.0	12391	8 ABZ69798	Abz69798 Plasmid P
5	459	22.3	4496	2 AAT58553	Aat58553 Streptomy
6	365.4	17.7	2888	2 AAT59268	Aat59268 Streptomy
7	361.8	17.6	2151	12 ADN36903	Adn36903 X. albifl
8	361.8	17.6	55839	12 ADN36893	Adn36893 X. albifl
9	316	15.3	26309	4 AAS59535	Aas59535 Propionib
10	316	15.3	26309	8 ACF64464	Acf64464 Propionib
11	270	13.1	1362	8 ACA42265	Aca42265 Prokaryot
12	268.4	13.0	1101	11 ABD14809	Abd14809 Pseudomon
13	268.4	13.0	1431	11 ABD15129	Abd15129 Pseudomon
14	267.2	13.0	2046	9 ADB06077	Adb06077 Altiococc
15	267.2	13.0	110000	9 ADB12064	ADB12064 Altiococc
16	266.8	12.9	1860	5 AAH66076	Aah66076 C glutam
17	266.8	12.9	1983	4 AAF72103	Aaf72103 Corynebac
18	266.8	12.9	1983	4 AAF71982	Aaf71982 Corynebac
19	266.8	12.9	2005	10 ADD13458	Add13458 C. glutam
20	266.8	12.9	349980	5 AAH66527	Aah66527 C glutam

21	260	12.6	816	11 ABD15339	Abd15339 Pseudomon
22	234.8	11.4	42000	3 AAA63349	Aaa63349 Streptomy
23	234.8	11.4	63164	3 AAA63348	Aaa63348 Streptomy
24	231.6	11.2	1338	8 ACA4234	Aca4234 Prokaryot
25	224.2	10.9	2523	8 ACA29955	Aca29955 Prokaryot
26	220.4	10.7	1257	8 ACA45986	Aca45986 Prokaryot
27	219	10.6	2019	10 ACF71773	Acf71773 Photorhab
28	219	10.6	110000	10 ACF67367	ACF67367 49
29	219	10.6	110000	10 ACF65387	ACF65387_1
30	210.2	10.2	2160	10 AAD50710	Aad50710 Rhodopsu
31	210.2	10.2	2160	10 AAD50709	Aad50709 Rhodopsu
32	210.2	10.2	2160	12 AD005494	Ado05494 Rhodopsu
33	210.2	10.2	2160	12 AD005544	Ado05544 Rhodopsu
34	210.2	10.2	2160	12 AD005501	Ado05501 R. palust
35	201.2	9.8	2199	12 AD005540	Ado05540 Azospiril
36	198.2	9.6	1353	8 ACA36033	Aca36033 Prokaryot
37	197.8	9.6	1506	11 ABD02207	Abd02207 Pseudomon
38	196.6	9.5	1518	11 ABD01104	Abd01104 Klebsiell
39	192.8	9.4	349980	6 ABQ81845	Abq81845 Bifidobac
40	190.4	9.2	1323	8 ACA31740	Aca31740 Prokaryot
41	189.6	9.2	2190	10 AAD50708	Aad50708 A. tumefa
42	189.6	9.2	2190	12 AD005493	Ado05493 A. tumefa
43	184	8.9	110000	11 ADM27081	ADM27081_03
44	184	8.9	110000	11 ADM27081_04	ADM27081_04
45	182.4	8.9	110000	4 AAI99682_17	AAI99682_17

ALIGNMENTS

RESULT 1	AAf66409	standard; DNA; 2061 BP.
ID	AAf66409	
XX	AAf66409;	
AC	22-JUN-2001	(first entry)
XX		
DE	4-amino-4-deoxychorismic acid synthase, papa, coding sequence.	
XX		
KM	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;	
KW	4-amino-4-deoxychorismic acid synthase; enzyme; papa; ds.	
XX		
OS	Streptomyces venezuelae.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..2061
FT		/tag= a
XX		/product= "4-amino-4-deoxychorismic acid synthase"
XX		
PN	WO200123542-A1.	
XX		
PD	05-APR-2001.	
XX		
PF	29-SEP-2000; 2000MO-JP06783.	
XX		
PR	29-SEP-1999; 99JP-00276314.	
XX		
PA	(MEIJ) MEIJI SEIKA KAISHA LTD.	
XX		
PI	Yanai K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Midoh N;	
PI	Murakami T;	
XX		
DR	WPI; 2001-290517/30.	
DR	P-PDB; AAB82071.	
XX		
PT	Transformant producing secondary metabolite modified with functional	
PT	group e.g. benzene with nitrogen-containing substituent at para-position,	
PT	pp1022, with ease at low cost, for application in pharmaceuticals and	
XX	agrochemicals.	
PS	Claim 11; Page 54-59; 83pp; Japanese.	
XX		

CC The present invention relates to a transformant having been modified so
CC as to produce a secondary metabolite. The secondary metabolite has a
CC benzene ring skeleton free from substitution at the para-position by a
CC nitrogen-containing functional group, thereby enabling the production of
CC a secondary metabolite with a benzene ring skeleton substituted at the
CC para-position by a nitrogen-containing group. The transformant organism
CC of the present invention has been produced by transferring a gene
CC participating in the biosynthesis pathway from chorismic acid into p-
CC anthonylpyruvic acid. The present sequence is the coding sequence for
CC 4-amino-4-deoxychorismic acid synthase (papa), from *Streptomyces*
CC *venezuelae*. papa participates in the biosynthesis pathway from chorismic
CC acid into p-anthonylpyruvic acid, and so the papa gene can be used to
CC produce the transformant of the present invention. The transformant can
CC be used to produce metabolites for application in pharmaceuticals,
CC veterinary drugs and agrochemicals

XX
SO Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other:

Query Match 100.0%; Score 2061; DB 4; Length 2061;
Best Local Similarity 100.0%; Pred. No. 1.7e-253;
Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGCGACGCTTGTGATCGACAACTAGACTGTTACCCCAACCTGTTCCAGTACATC 60
DB 1 ATGCGACGCTTGTGATCGACAACTAGACTGTTACCCCAACCTGTTCCAGTACATC 60
QY 61 GCGGAGGCCACCGGAGCAACCCCGTGTGTGTCGCCACGACGCCGAGTGTGTGCGGCTG 120
DB 61 GCGGAGGCCACCGGAGCAACCCCGTGTGTGTCGCCACGACGCCGAGTGTGTGCGGCTG 120
QY 121 CCGGTGAGGACTTTCGACGCGGATGTGTGTGTCGCCGAGCCCGGACGCCGAGCCGAGAA 180
DB 121 CCGGTGAGGACTTTCGACGCGGATGTGTGTGTCGCCGAGCCCGGACGCCGAGCCGAGAA 180
QY 181 CCGGACTTTCGGAATTCAGCCGCCCGGGGCGATATCCGACAGCGCCCTGCTCTGCGCTC 240
DB 181 CCGGACTTTCGGAATTCAGCCGCCCGGGGCGATATCCGACAGCGCCCTGCTCTGCGCTC 240
QY 241 TGCTGTGCGGACCAAGGAGCATGCGCAGCTCTTGTGCGGAGAACGTTGCGCTTCGCGGAA 300
DB 241 TGCTGTGCGGACCAAGGAGCATGCGCAGCTCTTGTGCGGAGAACGTTGCGCTTCGCGGAA 300
QY 301 CCGATGACGCGCGGGGTCTCCAGAGTGCAGCACCGCGGAGGACGTTCTCCGGGGCTTC 360
DB 301 CCGATGACGCGCGGGGTCTCCAGAGTGCAGCACCGCGGAGGACGTTCTCCGGGGCTTC 360
QY 361 CCTTGTGCGCTTACCGCGCGGTGCTACACTCCCTGCGCGCACCGACCTTCCCGACGAG 420
DB 361 CCTTGTGCGCTTACCGCGCGGTGCTACACTCCCTGCGCGCACCGACCTTCCCGACGAG 420
QY 421 CTGGAACCCCTGCGCTGAGAGCGACGAGGGGTGTCATGAGGCTGCGGACCGCGAGAG 480
DB 421 CTGGAACCCCTGCGCTGAGAGCGACGAGGGGTGTCATGAGGCTGCGGACCGCGAGAG 480
QY 481 CCGCTGTGAGGCGTTCAGTTTCAACCGAGATTCATCGGACGACGACTTCCGCGGAGATC 540
DB 481 CCGCTGTGAGGCGTTCAGTTTCAACCGAGATTCATCGGACGACGACTTCCGCGGAGATC 540
QY 541 ATGCGCAACTTCCGCGACCTGCGCTGCGCCACACCGGAGCGACGCGGCGCGAC 600
DB 541 ATGCGCAACTTCCGCGACCTGCGCTGCGCCACACCGGAGCGACGCGGCGCGAC 600
QY 601 TCCCGTGAAGAACTCCACGTCGCGCGGTGAGCTGTGCGGACCGCGAAGAGTACGC 660
DB 601 TCCCGTGAAGAACTCCACGTCGCGCGGTGAGCTGTGCGGACCGCGAAGAGTACGC 660
QY 661 CCGGCTGTGCTGCTCCCGGAGGAGCACACGTTTGTGCTGAGACAGCTCCGCTCTCGAA 720
DB 661 CCGGCTGTGCTGCTCCCGGAGGAGCACACGTTTGTGCTGAGACAGCTCCGCTCTCGAA 720
QY 721 GCGGCTGTGCTGCTCTCTCTCTGCGGAGAGCAGCCGCGGCTGCGCGAGTACTTCACC 780
DB 721 GCGGCTGTGCTGCTCTCTCTCTCTGCGGAGAGCAGCCGCGGCTGCGCGAGTACTTCACC 780
```

```
QY 781 TACCGGTGCGCGACCGGCTGCTCTCCGTCCGCGGTCCGACGCGACCAACCGGAGCG 840
DB 781 TACCGGTGCGCGACCGGCTGCTCTCCGTCCGCGGTCCGACGCGACCAACCGGAGCG 840
QY 841 CCGCGCCCTTCTTCAACTGAGAGAGCAGCTGAAACCGCGAGGTCCTGCTGCC 900
DB 841 CCGCGCCCTTCTTCAACTGAGAGAGCAGCTGAAACCGCGAGGTCCTGCTGCC 900
QY 901 CCGGAATGCGCTTTCAGATTCAACCTGCGCTAGTGTGCTACTTGGCTTACGAGCTGAAG 960
DB 901 CCGGAATGCGCTTTCAGATTCAACCTGCGCTAGTGTGCTACTTGGCTTACGAGCTGAAG 960
QY 961 GCGGAGACCAACCGGAGACCGCGGACCGGTCGCCACCGGACCGCGGCTTCTCTTC 1020
DB 961 GCGGAGACCAACCGGAGACCGCGGACCGGTCGCCACCGGACCGCGGCTTCTCTTC 1020
QY 1021 GCCGACCGGCGCAATGCGCTTTCGACACACAGAAAGCTGCTACTCTGTGCGCTTCGAC 1080
DB 1021 GCCGACCGGCGCAATGCGCTTTCGACACACAGAAAGCTGCTACTCTGTGCGCTTCGAC 1080
QY 1081 CCGCGGAGCGCACGAGACGAGCGCGCTGCTGCTGCGGAGACGCGCGGACCTTCACC 1140
DB 1081 CCGCGGAGCGCACGAGACGAGCGCGCTGCTGCTGCGGAGACGCGCGGACCTTCACC 1140
QY 1141 GCGCTGCGGCTGCGCGCCCGGCGGACCGGACCGGCGGATGCTTTCGAGATCCCGAG 1200
DB 1141 GCGCTGCGGCTGCGCGCCCGGCGGACCGGACCGGCGGATGCTTTCGAGATCCCGAG 1200
QY 1201 GCGGCGGCGGCTTTCGCGCCCTTTCGCGCGCGCGCGCACGACAGAGACGCTTACCTCAAG 1260
DB 1201 GCGGCGGCGGCTTTCGCGCCCTTTCGCGCGCGCGCGCACGACAGAGACGCTTACCTCAAG 1260
QY 1261 CCGATGAGAGAGTGTCTCAAGAGATTCGCGAAGCGGAGATCGTACGAGATTCGCTGACC 1320
DB 1261 CCGATGAGAGAGTGTCTCAAGAGATTCGCGAAGCGGAGATCGTACGAGATTCGCTGACC 1320
QY 1321 AACATGATCAACGCGGCGGACCGAGGCGAGCGGCGCTGCTCACTCCGCGCTGCGCGCC 1380
DB 1321 AACATGATCAACGCGGCGGACCGAGGCGAGCGGCGCTGCTCACTCCGCGCTGCGCGCC 1380
QY 1381 ATGAGCCCGCTCCGTAACGAGCGCCCTGCTCGAGTTCCCGGAACCTGCTGAGCGGCT 1440
DB 1381 ATGAGCCCGCTCCGTAACGAGCGCCCTGCTCGAGTTCCCGGAACCTGCTGAGCGGCT 1440
QY 1441 TCGCCGGAACGGTTCTTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG 1500
DB 1441 TCGCCGGAACGGTTCTTCAAGATCGGCGCGGACGCGGCGGCTGAGTCCAAAGCCATCAAG 1500
QY 1501 GGGACCCGCGCCCGGAGGAGCACCGGAGAGAGAGCAGCGGCTCCGCGCGGACCTGAGCC 1560
DB 1501 GGGACCCGCGCCCGGAGGAGCACCGGAGAGAGAGCAGCGGCTCCGCGCGGACCTGAGCC 1560
QY 1561 GCGCGGAGAGAGACCGGAGCGAGAACCTGATGATGCTGAGCTGTGCGAACGACCTTC 1620
DB 1561 GCGCGGAGAGAGACCGGAGCGAGAACCTGATGATGCTGAGCTGTGCGAACGACCTTC 1620
QY 1621 AACAGGCTGTGCGGATCGGCTTCGCTGACGTCGTCGCTCCGCTCTTGAAGTGAAGACTTAC 1680
DB 1621 AACAGGCTGTGCGGATCGGCTTCGCTGACGTCGTCGCTCCGCTCTTGAAGTGAAGACTTAC 1680
QY 1681 GCGCCCGTGAACAGCTGCTGTCGACATCGGAGGAGCGGCTGCGGCGCGGACCGACGAC 1740
DB 1681 GCGCCCGTGAACAGCTGCTGTCGACATCGGAGGAGCGGCTGCGGCGCGGACCGACGAC 1740
QY 1741 GCGGCTGTGCTGAGCGCGCTTTCGCGGCGGCTGATGACGCGCGCGGCTTCTTCAAGAGCGCC 1800
DB 1741 GCGGCTGTGCTGAGCGCGCTTTCGCGGCGGCTGATGACGCGCGCGGCTTCTTCAAGAGCGCC 1800
QY 1801 ACCATGAGATCATGACGCGCTTGAAGAGAGCGCCCGGAGCGTCTTACTTCGCGGCGCTTC 1860
DB 1801 ACCATGAGATCATGACGCGCTTGAAGAGAGCGCCCGGAGCGTCTTACTTCGCGGCGCTTC 1860
```

QY 1861 GAGATGTTCCCTCAGCGCGCCGACCTCAGATTCATCCGACCATCTGCTG 1920
Db 1861 GAGATGTTCCCTCAGCGCGCCGACCTCAGATTCATCCGACCATCTGCTG 1920
QY 1921 GCCGACCGCCAGCGAGTTTCGGCTCGCGGGGCGCATCTGTTCTCTCCGACGAG 1980
Db 1921 GCCGACCGCCAGCGAGTTTCGGCTCGCGGGGCGCATCTGTTCTCTCCGACGAG 1980
QY 1981 GAGAGATTCAACGAGACCGTGTAAAGCCCGGCCCATGATCCGCTCGACGGGAGC 2040
Db 1981 GAGAGATTCAACGAGACCGTGTAAAGCCCGGCCCATGATCCGCTCGACGGGAGC 2040
QY 2041 GCCGTGCGGGCGCGCCGATGA 2061
Db 2041 GCCGTGCGGGCGCGCCGATGA 2061

RESULT 2
AAL50180
ID AAL50180 standard; DNA; 2061 BP.
XX
XX AAL50180;
AC
XX
XX 28-JAN-2003 (first entry)
DT
XX
XX S venezuelae PF1022 substance gene #1.
DE
XX
XX
XX Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;
KM phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
KW veterinary drug; gene; ds.
XX
XX Streptomyces venezuelae.
OS
XX
FH Key Location/Qualifiers
FT 1..2061
FT /*tag= a
FT /product= "PF1022 substance"
XX
XX
XX MO200277244-A1.
XX
XX 03-OCT-2002.
PD
XX
XX 22-MAR-2002; 2002MO-JP002782.
PF
XX
XX 22-MAR-2001; 2001JP-00082227.
PR
XX
XX (MEIJ) MEIJI SEIKA KAISHA LTD.
PA
XX
XX Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;
PI
XX
XX WPI; 2003-018934/01.
DR
XX
XX P-PSDB; AAO19563.
DR
XX
XX Novel biosynthesis gene-transferred transformants for producing PF1022
PT substance derivatives by fermentation, as pharmaceuticals or veterinary
PT drugs with anthelmintic activity.
XX
XX
XX Claim 13; Page 74-79; 116pp; Japanese.
XX
XX The present invention relates to transformants capable of producing
CC PF1022 substance derivatives. These were obtained by transferring a gene
CC participating in the biosynthesis pathway from chorismic acid to p-
CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-
CC requiring host derived from an organism producing the PF1022 substance.
CC The transformants are producing PF1022 substance derivatives by
CC fermentation, for use as pharmaceuticals or veterinary drugs. The present
CC sequence is a substance PF1022 coding sequence from Streptomyces
CC venezuelae
XX
XX Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other;
SQ

Query Match 100.0%; Score 2061; DB 8; Length 2061;
Best Local Similarity 100.0%; Pred. No. 1.7e-253;

Matches 2061; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGGACACGCTTTCATCCGACCACTACGATCGTTTCAACCCACAACTGTTCCGATATC 60
Db 1 ATGGGACACGCTTTCATCCGACCACTACGATCGTTTCAACCCACAACTGTTCCGATATC 60
QY 61 GGGAGGCGCACCGGGGAAACCCCGGTCGTGTCGCCAACGACCGCGACTGTGCGCGCTG 120
Db 61 GGGAGGCGCACCGGGGAAACCCCGGTCGTGTCGCCAACGACCGCGACTGTGCGCGCTG 120
QY 121 CCCGTGAGGACTTCGACGGGATGTTGTGTCTCCCGGGCCCGGCAAGCCCGGAA 180
Db 121 CCCGTGAGGACTTCGACGGGATGTTGTGTCTCCCGGGCCCGGCAAGCCCGGAA 180
QY 181 CGGAGATTCCGGAATTCAGCCGCCCGGGGATACCCGACAGCGGCTTGCCCTGCGGCTC 240
Db 181 CGGAGATTCCGGAATTCAGCCGCCCGGGGATACCCGACAGCGGCTTGCCCTGCGGCTC 240
QY 241 TGCCTGAGCCACAGGAGCATCGCCAGCTCTTCGAGCGGAAACCGTCCGCTCGCCCGGAA 300
Db 241 TGCCTGAGCCACAGGAGCATCGCCAGCTCTTCGAGCGGAAACCGTCCGCTCGCCCGGAA 300
QY 301 CCCATGACAGCGCCGGGTCTCCAGGTGCGGCAACCGCGAGGACGTTCTCCGGGCGCTC 360
Db 301 CCCATGACAGCGCCGGGTCTCCAGGTGCGGCAACCGCGAGGACGTTCTCCGGGCGCTC 360
QY 361 CCTTCGCGGTTTCAACCGCCGTCGCTTACACTCCCTTGGCCGCGACCGACTCTCCCGACGAG 420
Db 361 CCTTCGCGGTTTCAACCGCCGTCGCTTACACTCCCTTGGCCGCGACCGACTCTCCCGACGAG 420
QY 421 CTGGAACCCCTCGCTGAGCGGACGAGCGGGGTGTATGAGGCTTGAGGACCGCGAGAG 480
Db 421 CTGGAACCCCTCGCTGAGCGGACGAGCGGGGTGTATGAGGCTTGAGGACCGCGAGAG 480
QY 481 CCGCTGTGGGCGCTTCAGTTTCCACCCGAGTCCATCGGACGACTTTCGCGCGGAGATC 540
Db 481 CCGCTGTGGGCGCTTCAGTTTCCACCCGAGTCCATCGGACGACTTTCGCGCGGAGATC 540
QY 541 ATGGCCAACTTTCGCGACACTTCGCTTCGCCACACCGGCGACAGCGGAGCGAC 600
Db 541 ATGGCCAACTTTCGCGACACTTCGCTTCGCCACACCGGCGACAGCGGAGCGAC 600
QY 601 TCCCGTAAAGAACTTCACAGCGCGCGGTCGAGTGTCTGCGGACCGCGAAGAGTAAAGC 660
Db 601 TCCCGTAAAGAACTTCACAGCGCGCGGTCGAGTGTCTGCGGACCGCGAAGAGTAAAGC 660
QY 661 CGCGGCTGCTGCGCGGACAGGACCAAGTTCGTGCTGAGACGAGCTTCGTTCTGAA 720
Db 661 CGCGGCTGCTGCGCGGACAGGACCAAGTTCGTGCTGAGACGAGCTTCGTTCTGAA 720
QY 721 GCGGCTCGGCTTCTCTTCTCTGCGGACGACCGGCGCCGTCGCGAGTACTTAC 780
Db 721 GCGGCTCGGCTTCTCTTCTCTGCGGACGACCGGCGCCGTCGCGAGTACTTAC 780
QY 781 TACCGGATGCGGACGCGCGCTGCTCCGTCGCGGACGAGCGCGGCTCGGACTTCAAC 840
Db 781 TACCGGATGCGGACGCGCGCTGCTCCGTCGCGGACGAGCGCGGCTCGGACTTCAAC 840
QY 841 CGCGCGCCCTTCTTCAACTACCTGAGAGCAGCTGAAACGCGAGCGGATCCCGTGC 900
Db 841 CGCGCGCCCTTCTTCAACTACCTGAGAGCAGCTGAAACGCGAGCGGATCCCGTGC 900
QY 901 CCCGAATCGCTTTCGAGTTCAACTTCGCGTACGTCGCTTACCTCGGCTTACGAGCTGAAG 960
Db 901 CCCGAATCGCTTTCGAGTTCAACTTCGCGTACGTCGCTTACCTCGGCTTACGAGCTGAAG 960
QY 961 GCGGAGACCAACGCGGACCCCGGACCGGATCCCGGACCCGAGCGCGGCTTCTTC 1020
Db 961 GCGGAGACCAACGCGGACCCCGGACCGGATCCCGGACCCGAGCGCGGCTTCTTC 1020
QY 1021 GCCGACCGGCGCATCGCTTCGACCAACGAGAGGCTGTGCTACTCTGCGCTTCGAC 1080
Db 1021 GCCGACCGGCGCATCGCTTCGACCAACGAGAGGCTGTGCTACTCTGCGCTTCGAC 1080

```
QY 1081 CGCCGGGGCCACGACGACGCGCCGCTGTGCGGAGACGCGCCGAGACCTTCACC 1140
DB 1081 CGCCGGGGCCACGACGACGCGCCGCTGTGCGGAGACGCGCCGAGACCTTCACC 1140
QY 1141 GGCCTGGCGCTCCGCGCCCGCGCCGAGCCGACCCCGCCATGCTTCCGGATCCCGAG 1200
DB 1141 GGCCTGGCGCTCCGCGCCCGCGCCGAGCCGACCCCGCCATGCTTCCGGATCCCGAG 1200
QY 1201 GCGCGCGCGCGCTTCCGCGCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
DB 1201 GCGCGCGCGCGCTTCCGCGCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 1261 CGCATCGACGAGTGCCTCAAGAGAGATCCGCAACGCGCGAGTCTGACAGATCTGCGAC 1320
DB 1261 CGCATCGACGAGTGCCTCAAGAGAGATCCGCAACGCGCGAGTCTGACAGATCTGCGAC 1320
QY 1321 AACATGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
DB 1321 AACATGCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380
QY 1381 ATCAGCGCGCTCCGCGTACGCGCGCGCGCTGAGTTCGCGACGCTGCGAGCGCGC 1440
DB 1381 ATCAGCGCGCTCCGCGTACGCGCGCGCGCTGAGTTCGCGACGCTGCGAGCGCGC 1440
QY 1441 TCGCCCGAGCGGCTTCTCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
DB 1441 TCGCCCGAGCGGCTTCTCAAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
DB 1501 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
QY 1561 GGGCGGGAGAGAGACGCGCGCGAGAACTGATGCTGCGACGCGCGCGCGCGCGCG 1620
DB 1561 GGGCGGGAGAGAGACGCGCGCGAGAACTGATGCTGCGACGCGCGCGCGCGCGCG 1620
QY 1621 AACAGCGTCTGCGCGCTGCTCAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
DB 1621 AACAGCGTCTGCGCGCTGCTCAAGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
QY 1681 GCGCGCGTGCACACGCTGCTGACACCATCCGCGGACGCGCTGCGCGCGCGCG 1740
DB 1681 GCGCGCGTGCACACGCTGCTGACACCATCCGCGGACGCGCTGCGCGCGCGCG 1740
QY 1741 GCGCGCGTGCACACGCTGCTGACACCATCCGCGGACGCGCTGCGCGCGCGCG 1800
DB 1741 GCGCGCGTGCACACGCTGCTGACACCATCCGCGGACGCGCTGCGCGCGCGCG 1800
QY 1801 ACCATGAGATCACTGACGCGCTGAGAGAGAGCGCGCGCGCGCGCGCGCGCG 1860
DB 1801 ACCATGAGATCACTGACGCGCTGAGAGAGAGCGCGCGCGCGCGCGCGCGCG 1860
QY 1861 GATGCTGCTGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
DB 1861 GATGCTGCTGCGCTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
QY 1921 GCGGACGCGCGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
DB 1921 GCGGACGCGCGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1980
QY 1981 GAGAGATTCAACGAGACCGTGTAAAGAGCGCGCGCGCGCGCGCGCGCGCG 2040
DB 1981 GAGAGATTCAACGAGACCGTGTAAAGAGCGCGCGCGCGCGCGCGCGCGCG 2040
QY 2041 GCCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061
DB 2041 GCCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061
```

RESULT 3
AB269799
ID AB269799 standard; DNA: 3305 BP.

```
XX AC AB269799;
XX DT 08-APR-2003 (first entry)
XX DE Plasmid papABC.
XX KM DHFR; translation; orthogonal RNA; O-tRNA; O-RS; CAT; TyrRS;
XX KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
XX KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
XX KM biosynthesis; p-aminophenylalanine; pAF; chorismate.
OS Synthetic.
PN WO200285923-A2.
PD 31-OCT-2002.
PF 19-APR-2002; 2002MO-US012465.
PR 19-APR-2001; 2001US-0285030P.
PR 06-FEB-2002; 2002US-0355514P.
XX (Scrip ) SCRIPPS RES INST.
XX Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TU,
PI Meggers EL, Mehli RA, Pasternak M, Santoro SW, Zhang Z;
XX WPI; 2003-120430/11.
XX
XX PT Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
XX PS Example 4; Page 127-128; 188pp; English.
XX
XX CC The invention relates to a novel composition comprising a translation
XX CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
XX CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
XX CC -tRNA with at least one unnatural amino acid in the translation system
XX CC and the O-tRNA recognises at least one selector codon. A composition of
XX CC the invention is useful for producing at least one protein comprising at
XX CC least one unnatural amino acid. The protein is the Asp127G mutant of
XX CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
XX CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
XX CC dihydrofolate reductase comprising CHO08 His6tag. The unnatural amino
XX CC acid is provided exogenously. The translation system is a cell and the
XX CC unnatural amino acid is biosynthesised by the cell. The present sequence
XX CC represents a plasmid containing the individual genes papABC that encode
XX CC the enzymes used to carry out the conversion of chorismate to the
XX CC unnatural amino acid p-aminophenylalanine (pAF)
XX
XX SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;
XX
XX Query Match 96.1%; Score 1773.8; DB 8; Length 3305;
XX Best Local Similarity 92.8%; Pred. No. 4.5e-217;
XX Matches 1913; Conservative 0; Mismatches 97; Indels 51; Gaps 3;
QY 1 ATGCGACCGCTTGTGATCGACAACTGACGACTGTTCAACCCACAACTGTTCCAGTACATC 60
DB 1 ATGCGACCGCTTGTGATCGACAACTGACGACTGTTCAACCCACAACTGTTCCAGTACATC 60
QY 61 GGGAGAGCCACCGGGAGACCCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
DB 61 GGGAGAGCCACCGGGAGACCCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 120
QY 61 GGGAGAGCCACCGGGAGACCCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
DB 61 GGGAGAGCCACCGGGAGACCCCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCG 117
QY 121 CCCGTGAGAGACTTTCAGACGCGATCGTGTGCTCCCGCGCGCGCGCGCGCGCGCG 180
DB 121 CCCGTGAGAGACTTTCAGACGCGATCGTGTGCTCCCGCGCGCGCGCGCGCGCGCG 180
QY 118 CCCCTGAGAGACTTTCAGACGCGATCGTGTGCTCCCGCGCGCGCGCGCGCGCG 177
DB 118 CCCCTGAGAGACTTTCAGACGCGATCGTGTGCTCCCGCGCGCGCGCGCGCGCG 177
QY 181 CGGAGCTTTCAGATACGCGCGCGCGCGCGATACCGACGCGCGCGCGCGCGCGCG 240
DB 178 CGGAGCTTTCAGATACGCGCGCGCGCGATACCGACGCGCGCGCGCGCGCGCG 237
```



```

QY 241 TGCTCGGACCAACAGGCGATCGCCAGCTCTTGGCGGAAACCGTGGGCTGCCCCGAAA 300
Db 238 TGCTCGGACCAACAGGCGATCGCCAGCTCT-----CGCGGAAA 276
QY 301 CCATGACGAGCGGAGTCTCCGAGTGCAGCAACGAGAGAGATCTTCCGAGGCTC 360
Db 277 CCATGACGAGCGGAGTCTCCGAGTGCAGCAACGAGAGAGATCTTCCGAGGCTC 336
QY 361 CCCCTGCGCTTCAACCGCGCTGCGCTACCACTCCCTGGCCGACCGACTTCCCGACGAG 420
Db 337 CCCCTGCGCTTCAACCGCGCTGCGCTACCACTCCCTGGCCGACCGACTTCCCGACGAG 396
QY 421 CTGAAACCCCTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db 397 CTGAAACCCCTCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 456
QY 481 CGGCTGAGGAGCGCTCAAGTTCACACCGAGTCAATCGGACGACTTCCGCGGAGATC 540
Db 457 CGGCTGAGGAGCGCTCAAGTTCACACCGAGTCAATCGGACGACTTCCGCGGAGATC 516
QY 541 ATGAGCAACTTCCGCGACCTCGCCCTGCGCCACACCGAGGACGCTCGCGAGAC 600
Db 517 ATGAGCAACTTCCGCGACCTCGCCCTGCGCCACACCGAGGACGCTCGCGAGAC 576
QY 601 TCCCGTACGAACTCGACGTCGCGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 660
Db 577 TGGAGCTACGAACTCGACGTCGCGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 636
QY 661 CGGCGCTGCTGCGCGCGAGGAGACGACTTGGCTGAGACAGACGCTCGCTCTCGAA 720
Db 637 CGGCGCTGCTGCGCGCGAGGAGACGACTTGGCTGAGACAGACGCTCGCTCTCGAA 696
QY 721 GGGGCTCGGCTTCTCTCTCTCGGAGACGACGCGCGCGCTCGCGAGTACTCACC 780
Db 697 GGGGCTCGGCTTCTCTCTCTCGGAGACGACGCGCGCGCTCGCGAGTACTCACC 756
QY 781 TACCGGCTGCGAGCGGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 840
Db 757 TACCGGCTGCGAGCGGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 816
QY 841 CGGCGCGCTTCTTCACTACCTGAGAGAGAGTCAAGACGCGAGCGGCTCGCGTGC 900
Db 817 CGGCGCGCTTCTTCACTACCTGAGAGAGAGTCAAGACGCGAGCGGCTCGCGTGC 876
QY 901 CGGCACTGCGCTTCACTACCTGAGTCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 960
Db 877 CGGCACTGCGCTTCACTACCTGAGTCAAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 936
QY 961 GGGGAGACACCGGCGAGCGCGCGAGTCCCGGACCGCGAGCGCGGCTCTCTTC 1020
Db 937 GGGGAGACACCGGCGAGCGCGCGAGTCCCGGACCGCGAGCGCGGCTCTCTTC 996
QY 1021 GCGGACGCGCGCATCGGCTTCACTACCAAGAGAGTGTACTGCTGCGCTCGAC 1080
Db 997 GCGGACGCGCGCATCGGCTTCACTACCAAGAGAGTGTACTGCTGCGCTCGAC 1056
QY 1081 CGCGGAGGACACGAGCGGCGCGCGCTGCTGCGGAGACGCGCGAGACCTTCACC 1140
Db 1057 CGCGGAGGACACGAGCGGCGCGCGCTGCTGCGGAGACGCGCGAGACCTTCACC 1116
QY 1141 GGGCTGCGCGCTCGCGCGCGCGCGAGCGACCGCGCGCATGCTTCCGAGATCCCGAG 1200
Db 1117 GGGCTGCGCGCTCGCGCGCGCGCGAGCGACCGCGCGCATGCTTCCGAGATCCCGAG 1176
QY 1201 GCGGCGCGCGCTTCCGCGCGCTGAGCGCGCGCGCGAGACGAGTGCAGTGCAG 1260
Db 1177 GCGGCGCGCGCTTCCGCGCGCTGAGCGCGCGCGCGAGACGAGTGCAGTGCAGTGCAG 1228
QY 1261 CGCATGACGAGTGCATCAAGAGATCGCAACGCGCGAGTGCAGTGCAGTGCAGTGCAG 1320
Db 1229 -----CGGCGCTCGCAACGCGCGAGTGCAGTGCAGTGCAGTGCAGTGCAGTGCAG 1269

```

```

QY 1321 AACATGATCACCGGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1380
Db 1270 AACATGATCACCGGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1329
QY 1381 ATGAGCCCGCTCCGATACCGGCGCGCTGCTGAGTTCGCCAATGCTGAGTGCAGCGCC 1440
Db 1330 ATGAGCCCGCTCCGATACCGGCGCGCTGCTGAGTTCGCCAATGCTGAGTGCAGCGCC 1389
QY 1441 TCGCGGAGCGGCTTCTTCACTACGATCGGCGCGAGCGCGCGCTGATGATCAAGCCCATAG 1500
Db 1390 TCGCGGAGCGGCTTCTTCACTACGATCGGCGCGAGCGCGCGCTGATGATCAAGCCCATAG 1449
QY 1501 GGGACCGCGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1560
Db 1450 GGGACCGCGCGCGCGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1509
QY 1561 GGGCGGAGAGAGACCGGCGCGAGAACTGATGATCTGACCTGATCGGCGAGCGAGCGAGCGAG 1620
Db 1510 GGGCGGAGAGAGACCGGCGCGAGAACTGATGATCTGACCTGATGATCGGCGAGCGAGCGAG 1569
QY 1621 AACAGGCTTCCGCGATCGGCTCCGTCACGTGCGCGCGCTTCTTCAAGTGCAGACTTAC 1680
Db 1570 AACAGGCTTCCGCGATCGGCTCCGTCACGTGCGCGCGCTTCTTCAAGTGCAGACTTAC 1629
QY 1681 GCGCGGTCACGAGCTGATGATCGAGCATCGGAGGAGCGAGTGCAGCGCGCGCGAGCGAGCGAG 1740
Db 1630 GCGCGGTCACGAGCTGATGATCGAGCATCGGAGGAGCGAGTGCAGCGCGCGCGAGCGAGCGAG 1689
QY 1741 GCGCGCTGCTGACGCGCGCGCTTCCCGCGCGCTGATGACCGGCGCGCGCGCGAGCGAGCGAG 1800
Db 1690 GCGCGCTGCTGACGCGCGCGCTTCCCGCGCGCTGATGACCGGCGCGCGCGAGCGAGCGAG 1749
QY 1801 ACCATGAGATCATCGACGCGCTGAGAGAGCGCGCGCGCGCTTCTTCAAGTGCAGCGCTC 1860
Db 1750 ACCATGAGATCATCGACGCGCTGAGAGAGCGCGCGCGCGCTTCTTCAAGTGCAGCGCTC 1809
QY 1861 GGAATGTTGCGCTTCAAGCGCGCGCGCGAGCTGAGATGCTATCGGACCATCGTCTG 1920
Db 1810 GGAATGTTGCGCTTCAAGCGCGCGCGCGAGCTGAGATGCTATCGGACCATCGTCTG 1869
QY 1921 GCGGAGCGCGAGCGAGTTCGCGCTGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1980
Db 1870 GCGGAGCGCGAGCGAGTTCGCGCTGCGCGAGCGAGCGAGCGAGCGAGCGAGCGAGCGAG 1929
QY 1981 GAGGAGTTCAACGAGACCGTGTAAAGCGCGCGCGCATGCTACCGCGCTGACGCGAGC 2040
Db 1930 GAGGAGTTCAACGAGACCGTGTAAAGCGCGCGCGCATGCTACCGCGCTGACGCGAGC 1989
QY 2041 GCGGTGCGGCGCGCGCGATGA 2061
Db 1990 GCAGTGCGGCGCGCGCGATGA 2010

RESULT 4
ABZ69798
ID ABZ69798 standard; DNA; 12391 BP.
XX
AC ABZ69798;
XX
DT 08-APR-2003 (first entry)
XX
DE plasmid pasc-papabc.
XX
KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyRS;
XX orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
XX chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
XX biosynthesis; p-antimphenylalanine; paf.
OS Synthetic.
XX
PN MO200285923-A2.
XX
PD 31-OCT-2002.

```

XX 19-APR-2002; 2002MO-US012465.
XX
XX 19-APR-2001; 2001US-0285030P.
PR 06-FEB-2002; 2002US-0355514P.
XX
XX (SCRI) SCRIPPS RES INST.
XX
PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ,
PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
XX
XX WPI; 2003-120430/11.
XX
XX
XX Composition useful for producing protein comprising unnatural amino acid,
XX has translation system comprising orthogonal tRNA and orthogonal
XX aminoacyl tRNA synthetase.
XX
XX
XX Example 4; Page 124-127; 188pp; English.
XX
XX The invention relates to a novel composition comprising a translation
XX system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
XX tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
XX -tRNA with at least one unnatural amino acid in the translation system
XX and the O-tRNA recognizes at least one selector codon. A composition of
XX the invention is useful for producing at least one protein comprising at
XX least one unnatural amino acid. The protein is the Asp12TAG mutant of
XX chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
XX dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
XX acid is provided exogenously. The translation system is a cell and the
XX unnatural amino acid is biosynthesized by the cell. The present sequence
XX represents a plasmid for use in the biosynthesis of p-aminophenylalanine
XX (PAF) in vivo
XX
XX Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;
XX
XX Query Match 86.0%; Score 1772.2; DB 8; Length 12391;
XX Best Local Similarity 92.8%; Pred. No. 5.5e-217;
XX Matches 1912; Conservative 0; Mismatches 98; Indels 51; Gaps 3;
XX
QY 1 ATGCGACGCTTCTGATCGACAACTAGACTCGTTCCACCAACCTGTTCCAGTACATC 60
DB 24 ATGCGACGCTTCTGATCGACAACTAGACTCGTTCCACCAACCTGTTCCAGTACATC 83
QY 61 GGGAGGCGACCGGCGCAACCCCGTGTGTCGTCGCCAAGCGCGCACTGTCGCGGCTG 120
DB 84 GGGAGGCGACCGGCGCAACCCCGTGTGTCGTCGCCAAGCGCGCACTGTCGCGGCTG 140
QY 121 CCGGTGAGGACTTTCGACGCGATCGTGTGTCGCCGCGGCGCGCGACCGCGCAACCGGAA 180
DB 141 CCGGTGAGGACTTTCGACGCGATCGTGTGTCGCCGCGGCGCGCGACCGCGCAACCGGAA 200
QY 181 CCGGACCTTCGGAATCAGCCCGCGGCGATCAGCGACGCGGCTGCGGCTTCCTCGGCTC 240
DB 201 CCGGACCTTCGGAATCAGCCCGCGGCGATCAGCGACGCGGCTGCGGCTTCCTCGGCTC 260
QY 241 TGCTCGGCGACCGAGGCGATCGCCAGCTCTTGCGGCGGAACCGTGGCTCGGCCGGA 300
DB 261 TGCTCGGCGACCGAGGCGATCGCCAGCTCT-----CGGCGGA 299
QY 301 CCGATGACGCGCGGCTCTCCAGGTGCGGACACCGGCGAGGACGTCTTCGCGGCGCTC 360
DB 300 CCGATGACGCGCGGCTCTCCAGGTGCGGACACCGGCGAGGACGTCTTCGCGGCGCTC 359
QY 361 CCGTCCGCGTTTCAACCGCGGTGCGTACACTCCCTGGCGCGACCGACTTCCTCCGAGAG 420
DB 360 CCGTCCGCGTTTCAACCGCGGTGCGTACACTCCCTGGCGCGACCGACTTCCTCCGAGAG 419
QY 421 CTGGAACCTCTCGCTGGAAGCGACGAGGGGTGTATGAGGCTGCGGCGACCGCGAGAG 480
DB 420 CTGGAACCTCTCGCTGGAAGCGACGAGGGGTGTATGAGGCTGCGGCGACCGCGAGAG 479
QY 481 CCGCTGTGGGCGTCCAGTTCCACCGGAGTCCATCGGCGAGCGACTTCGCGCGGAGATC 540

DB 480 CCGCTATGGCGCTCCAGTTCCACCGGAGTCCATCGGACGCACTTCGCGCGGAGATC 539
QY 541 ATGCGCACTTCGCGGACCGCTCGCGCGACCAACCGGACGCGCGACCGGCGCGAC 600
DB 540 ATGCGCACTTCGCGGACCGCTCGCGCGACCAACCGGACGCGCGACCGGCGCGAC 599
QY 601 TCCCGCTACGAATCCACGTCGCGCGCGTGCAGCTGCTCGCGACCGCGAAGAGTACG 660
DB 600 TGGGGCTACGAATCCACGTCGCGCGCGTGCAGCTGCTCGCGACCGCGAAGAGTACG 659
QY 661 CCGGCTGCTGCGCGCGCGGCGGACCAAGTTCTGCTGGAACGACTTCCTTCGAA 720
DB 660 CCGGCTGCTGCGCGCGCGGCGGACCAAGTTCTGCTGGAACGACTTCCTTCGAA 719
QY 721 GCGCGTCCGCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 780
DB 720 GCGCGTCCGCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 779
QY 781 TACCGGTGCGCGACCGCGTGTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 840
DB 780 TACCGGTGCGCGACCGCGTGTCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 839
QY 841 CCGCGCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 900
DB 840 CCGCGCGCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTCCTTC 899
QY 901 CCGGACCTGCGCTTCGATTCACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960
DB 900 CCGGACCTGCGCTTCGATTCACACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 959
QY 961 GCGGAGACCAACCGGCGACCGCGCGACCGGTCGCCACCGCGCGCTTCCTTCCTTC 1020
DB 960 GCGGAGACCAACCGGCGACCGCGCGACCGGTCGCCACCGCGCGCTTCCTTCCTTC 1019
QY 1021 GCGGACCGGCGCATGCGCTTCGACACGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1020 GCGGACCGGCGCATGCGCTTCGACACGAGAAAGCTGCTGCTGCTGCTGCTGCTGCTG 1079
QY 1081 CCGCGGCGCGACGACGACCGCGCGCGCTGCTGCGGAGACCGCGCGACCGCTCAC 1140
DB 1080 CCGCGGCGCGACGACGACCGCGCGCGCTGCTGCGGAGACCGCGCGACCGCTCAC 1139
QY 1141 GCGCTGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200
DB 1140 GCGCTGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1199
QY 1201 GCGGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
DB 1200 GCGGCGCGCGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1251
QY 1261 CCGATGAGAGAGTCTCAAGAGATCGGCAACGCGGAGTCGTAGAGATTCGCGTAC 1320
DB 1252 -----CGGCGTTCGCAACGCGGAGTCTGATCGAATTCGCTGAC 1292
QY 1321 AACATGATCAACGCGCGCGACCGAGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1380
DB 1293 AACATGATCAACGCGCGCGACCGAGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1352
QY 1441 TCGCCGGAAGGCTTCCTACAGATCGCGCGCGAGCGGCGGTGAGATTCGAAGCCATCAAG 1500
DB 1413 TCGCCGGAAGGCTTCCTACAGATCGCGCGCGAGCGGCGGTGAGATTCGAAGCCATCAAG 1472
QY 1501 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560
DB 1473 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532
QY 1561 GCGCGGAGAGAGACCGGCGCGAGACCTGATGATGTCGACTGCTTCGCAACGACCTTC 1620


```

QY 476 AGAAGCCGCTGTGGGGGCTCCAGTTCACCCGGAGTCATGCGAGCGACTTCGGCCGG 535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3508 ACCCTGCCCTTCGGGTGCAAGTTCACCCGAATGATCAGACGAGCAACGGGCACC 3567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 536 AGATCATGGCCAACTTCGGGACCT----- 560
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3568 GGAATGCTGCGCAACTTCGGGACCTTCCTGTGGCGCGCGCGGACCGCCCCCGACA 3627
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 -----CGCCCTGCGCAACACCGGGGACGGCGGCGGCGGCGGCGGCGGCGGCGG 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3628 CCGAAGCGAATACCGGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3687
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 606 -----GTACGAATCCACGTGCGCGGCTGACAGTGTCCGAGC 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3688 CGTCCGCGCGGCTGGGGGAGTACCGGCTGATGTGGCGAGGTGCGCTGTGCCGACG 3747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 CCGAAGAGTACCGCGCGGCTGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 706
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3748 CCGAAGCGGCTTACCGGCTGTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3807
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 707 GCTCCGCTCTGAAAGCGCTCGCGCTTCTCTTCCTGCGGCGGCGGCGGCGGCGGCGG 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3808 GCGCGCTCGAGCGCGGCTGTGCGGCTTCACTTCTGCGGCGGCGGCGGCGGCGGCGG 3867
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 CCGAGTACTCTACCTACCGGCTGCGGCGGCGGCGGCTTCTGCGGCGGCGGCGGCGG 826
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3868 GCGAAGAGTACCTTACGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 827 CCAAGACCGGAGCGCGGCGGCGGCTTCTTCACTGAGGAGGAGGAGCTGGAACCGCGAG 886
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3928 GCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 887 GGGTCCCGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3985 GCGCGCTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 947 GCTACGAGCTGAAAGCGGAGACCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1006
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4045 GCTACGAGACCAAGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1007 CCGCGTTCCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1066
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4105 GCGCTTTCATGTTGCGCGGAGTGTGCGCTTCGACCAAGAGGCGGCGGCGGCGGCGG 4164
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1067 TGCTGCGCCTCGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1126
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4165 TCTTGGAAGTACGACGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4224
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1127 CCGAGACCTTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1186
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4225 CCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4284
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1187 TCGGAGTCCCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1246
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4285 AGCAACCACTGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4329
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1247 AGCGCTACCTCAAGCGCATGAGAGTCTCAAGAGATCCGCAACGGGCGGAGTCTGAG 1306
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4330 CCGGCTACCGGGAATCGTGTGAGGAATGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1307 AGATGCTGCTGCAACCATGTGTACCGCGCGGCGGCGGAGGCGGCGGCGGCGGCTTACT 1366
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4390 AGGTGTGCTGAGCAACATGCTCCGGGTGCGCGGCGGAGTCAACCGGCTACCGGCTTAC 4449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1367 CCGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4450 GCGCGCTGGCGACCGTGTAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6
 AAT59268
 ID AAT59268 standard; cDNA; 2888 BP.
 XX

```

AC AAT59268;
XX
XX 02-APR-1997 (first entry)
XX
DE Streptomyces pristinaespiralis papA and papM intergenic region.
XX
XX Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
XX DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
XX papC; isomerisation; aromatisation; N-methyltransferase; ds.
XX
OS Streptomyces pristinaespiralis.
XX
FH Key Location/Qualifiers
FH 1..687
FH FT /*tag= a
FH FT /product= "papA"
FH FT /note="C-terminal coding region only, i.e. a partial
FH FT open reading frame".
FH FT complement(949..1836)
FH CDS /*tag= b
FH FT /product= "PapC"
FH FT 1873..2262
FH FT /*tag= c
FH FT /product= "papB"
FH FT 2259..2888
FH FT /*tag= d
FH FT /product= "papM"
FH FT /note="N-terminal coding region only, i.e. a partial
FH FT open reading frame"
XX
XX MO9601901-A1.
XX
XX 25-JAN-1996.
XX
XX 04-UTL-1995; 95WO-FR00089.
XX
XX 08-UTL-1994; 94FR-00008478.
XX
XX (RHON ) RHONE POULENC RORER SA.
XX
XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,
XX Barriere J, Debussche L, Famechon A, Dutruc-Rosset G,
XX
XX WPI; 1996-097631/10.
XX P-PSDB; AAW11582, AAW11583.
XX
XX New streptogramin B derivs. useful as antibiotics - produced by new
XX mutants of Streptomyces having altered genes for streptogramin B
XX biosynthesis.
XX
XX Example 1; Page 102-104; 146pp; French.
XX
XX The papA and papM genes of S.pristinaespiralis are involved in the
XX biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
XX pristinamycin IA. The region between these two genes was sequenced and
XX two open reading frames were identified. The first (papC) was on the
XX complementary strand and decodes to an amino acid sequence with homology
XX to the region of E.coli TyrA which has been implicated in aromatisation
XX reactions. The papC gene product is likely to be involved in a similar
XX aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
XX phenylpyruvate during DMPAPA synthesis. The second open reading frame
XX (papB) could be decoded to give a product with homology to the region of
XX TyrA which has chorismate mutase activity. The papB gene product is
XX likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
XX deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
XX used to produce strains of S.pristinaespiralis which are unable to
XX produce the antibiotic pristinamycin I but which may be able to produce
XX new, modified forms of it
XX
XX
XX Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;
XX
XX Query Match 17.7%; Score 365.4; DB 2; Length 2888;
XX Best Local Similarity 74.2%; Pred. No. 3.1e-38;

```

	Matches	462;	Conservative	0;	Mismatches	161;	Indels	0;	Gaps	0;
Qy	1408	CTCGAGTTTCCCCGAACTGTCGGTGTGTTGAGCGCTTCGCCGACGGTTTCTTACGATCGGC								1467
Db	1	CTGCAAGTTCCTCCCGGAGCCACCGTGTCTCAGCTCTTCACTCCGAACGGTTCTCTCGATCGGC								60
Qy	1468	GCCGACGGCGGGGTCGTGAGTCCAAAGCCATCAAGGGAGACCGGCCCCCGGGGGGAGACCGCG								1527
Db	61	GCGAGCGCTGGCGGAGTCCAAACCATCAAGGGACCGCGCCCCCGGGGGGAGACCGGAGAC								120
Qy	1528	GAGANGACGAGCGGCTTCGCGCGGACCTTGGCCGGCGGGAGAAAGAACCGGGCGAGAAC								1587
Db	121	GCCGACGACCGCGCGCTTCAAGGCGTCTCCGCGCGGGCGAGAAAGAACCGGAGAGAAC								180
Qy	1588	CTGATGATCGTCGACTTGTTCGCAAGCACTTCAAAGCGTGTGCGGATGTGGCTCGTC								1647
Db	181	CTGATGATCGTCGACTTGTTCGCAAGCACTTGGCCAGGTCTGGACATCGGCTCGTTC								240
Qy	1648	CACGAGCCCGGGCTTGTGAGGTGAGACTAACGGCCCGGTCACAGCTGTGTGACCC								1707
Db	241	CACGTAACGGGCTGTTCGAGGTGAGACTTAACGCACTGTCACACACTTGTGACAGCG								300
Qy	1708	ATCCGGGACCGGCTTCGCGCCGGACACAGACCGCGGCTGTGTAACGGCGGCTTCCCT								1767
Db	301	GTCCGGCGCGGCTGTGGCGCGGACGTCTCCGCGCCCGCGGGGTACGGGCGGCTTCCCT								360
Qy	1768	GCGCGGCTTCATGACCGCGCGGCCCAAGAGCGCACATGAGATCATGACCGCTGGAG								1827
Db	361	GCGCGGCTTCATGACCGCGCGGCCCAAGGTCCGACACATGACATGATCATGACCGGCTGAG								420
Qy	1828	GAAAGCCCCCGGGGCGTCTACTTCCGGGGCGGCTCGATGTTCGCCCTTACGCGGCGGCC								1887
Db	421	AAGGACCCCGCGCGGTGTACTTCGAGCGCGGCTGTGAGCTACTTGCCCTTACGAGGCGCGCC								480
Qy	1888	GACCTTCAGCATCGTTCATCCGCAACCATCGTCTGCGCGACCGGCGGAGATTTCGAGCTC								1947
Db	481	GACCTTCAGCATCGTTCATCCGCAACCATCGTCTGCGCACGAGAGGCGGCCCAACCATCGGCTG								540
Qy	1948	GCGCGGCGGATCGTGTCTCTCTCTCCGACCGAGAGAGAGATTACCGAGAACCGTGTAAAG								2007
Db	541	GCGCGGCGGCGGCTGTGCGGCTGTCCGACCCGACGAGGATCGCGGAAATGCTCTCTAAG								600
Qy	2008	GCCCGCGGCATGCTTACCGGCTT								2030
Db	601	GCGCAGACCACTTCGCGGCTT								623
RESULT 7										
ADN36903										
ID	ADN36903	standard; DNA; 2151 BP.								
AC	ADN36903;									
XX										
DT	15-JUN-2004	(first entry)								
DE	X. albilineans XALB1 gene cluster DNA encoding protein AlbXVII.									
XX										
KW	AlbXVII family; antibiotic production; biosynthetic gene cluster.									
KW	XALB1; albicidin biosynthetic gene cluster 1; phytotoxic damage; AlbXVII;									
XX	gene; ds.									
OS	Xanthomonas albilineans.									
XX										
PN	MO2004035760-A2.									
PD	29-APR-2004.									

XX	PI	Royer M., Gabriel DM., Frutos R., Rott P.	
XX	DR	WPI; 2004-365158/34.	
XX	DR	P-PSDB; ADN36934.	
PT	XX	New transformed host cell, useful for producing antibiotics, preferably polyketide antibiotics for protecting plants against phytotoxic damage, or damage against abdicidin.	
PS	XX	Claim 8; SEQ ID NO 11; 193bp; English.	
SO	XX	Sequence 2151 BP; 418 A; 607 C; 665 G; 461 T; 0 U; 0 Other;	
Query Match	17.6%;	Score 361.8;	DB 12; Length 2151;
Best Local Similarity	51.4%;	Pred. No. 9.3e-38;	
Matches 1057;	Conservative 0;	Mismatches 932;	Indels 66; Gaps 7
QY	2	TGCGAGCGCTTGTGATGACAACTAGACATGACATCCCAACCGCTTCCAGTACATCG	61
DB	2	TGCGCTGCTTATCATTAATATATTACATTCGATTCCTTGAAATCTGCGCACTACGTAG	61
QY	62	GCGAGCGCACCGGGGCAACCCCGCTGCTGCTGCCCAACGACGCGCATGATGCGC----	116
DB	62	CGCAATCTTGGGGGAAGATCCCTGCTGATGCAACAACAGACAGATCTCTCGACAGAAC	121
QY	117	-GCTGCGCTGAGGACTTCGACGCGATCTGTGTTCCTCCGGGCCCCGACGCCGACC	175
DB	122	TGAAGAACCGGGGGGATTTTCCCTGATCATCTTTTCGCCGGCTCCGGCTCGGTGTTA	181
QY	176	GGGAACGGGAATTGGGAATACGCGCGGGGGGATACACAGCGGGCTGCCGCTCTCG	235
DB	182	ATGAAGCGGAATTTTCACTCTCGCTGAGGCGCTGAGACGAACGAATTTCCGATGTAG	241
QY	236	GCGTGTGCTGTGAGCACACGAGGAGCATCGCCAGCTCTTCCGGCGGAACCGTCCGCTCGCC	295
DB	242	GCGTATGCTGGCGCTTTCAGGGAATTGGGATGTCTATGTGTGGCCGATCTCGATGCC	301
QY	236	GCGAACCATCATCGGCGCGGGTCTCCGAGGTGGGCAACCGGCGAGGACGTCTTCGGG	355
DB	302	CGGTGCTCTTTCATGCGCGTGTCTCCACCGTCAATCAACCGGCGAGCGGTGTTGTTGAAG	361
QY	356	GCGTCTCCCTGCGCGTTTCAACGCGGTGTGCTTCACTTCCCTGCGCGCCACCGA-----	409
DB	362	GCATCTCCGCAACGTTTGGAGGCAATGTGCTATCATCTGTGATGTCTCCAGCAATGCG	421
QY	410	TCCCGGAGAGCTTGAACCCCTGCGCTCGAGGACAGACGAGGATCGTATGAGCTCGGCG	469
DB	422	TGCGGCTGTGTCTGAAGATGACGCGGCTTACCGATTTGCGGTGTGTGATGGCTTGGCAC	481
QY	470	ACCGGAGAACCGCTGTGGGCGGTCCAGTTTCCACCCGGAATTCATCGGACAGCATTTGG	529
DB	482	ACGTCAACACCCGAAATATGGGAGTACAGTTTCCACCCCGGAATGATCTTCACGAACAG	541
QY	530	GCGGGAATATATGGCCAACTTCCGGAAGCTCGCCCTCGCCCAACCGAGGACGCGCGC	589
DB	542	GCAACCGCATTTGTCTTAACTTTGCAACGATGCTTCCGCGCCACAGTGCACCGTTATCTTG	601
QY	590	ACGGGGCGGATCTCCCGTATGAACTCAGATGCGCGGCGTGCAGCGTGTGCGGAGCGCG	649
DB	602	CGGGGTGGAGACAGCGCGGCAAGTTTTTAAAGCTTTTGGCGCGCCGAGATGTGACACCGC	661
QY	650	AAGAGATACGCGCGGCTGCTGCGCGGCGAGGACACAGTTTCTGGCTGGA---CAGCA	706

```
Db      662  GGGTACGTGCGATGCTGAGCGCGGAAGATCAAGTGCCGTTGGCAGCGGAAGATGTCCTTTC 721
QY      707  GCTCCGTCCTCGGAAGGCGCTCCGCTTCTCTCTCTCCGCGAC-----GACCGCGGCC 760
Db      722  TGGCCTTGTTCGTGACGAAAGCATTTGCTTGGCTGACACAGCAGCTGGTCTGACATC 781
QY      761  CGCTCCCGGAGTACCTACCTTACCGCGTCCGCGAGCGCTGCTCTCGTCCGCGGCTCCG 820
Db      782  CAATGGCGCGCTTATGCTTATGAGACGGTGAACAGAGCGAGATGATGCGGATTTGCG 841
QY      821  ACGGACACAGACCCGAGCGCGCGCTTCTTCACTACCTGAGGA-----869
Db      842  TGGGCGACGAGGACATGATGACAGAGCAGCGAGCGTTCTTTCGTAAGATGATCGGG 901
QY      870  -GAGCTCGAAGCCCGACCGGCTCCCGTCCGCCCCCGAACTGCGCTTCAATTCACCTCG 928
Db      902  CGTTGCAATCGTGTCTTACTGAGGACGTCCGACGCGCCACCGTTCCGCTTTCGCGCG 961
QY      929  GCTACGTCGCTACTCTCGGCTACGAGCTGAAAGCGAGACACCGCGACCCCGCGCAC 988
Db      962  GCTACGTCGCTACTCTCGGCTACGAGCTGAAATGAAATCGGCTTCCGCGCGGCTTCA 1021
QY      989  GGTCCCGGACCCCGACGCGCGCTTCTTTCGCGACCGCGCATCGCTCGACCCACC 1048
Db      1022  CCAATGCAATCCCGATGCGCTTGTGATGCGCGTGAGCGCTTCTGTTGCTTGCACACG 1081
QY      1049  AGGAAGGCTGCTGCTACTCTGCTGCGCTCCGCTCGACCGCGGCGCACAGACGCGCGCG 1108
Db      1082  CCACTGAGAGGATATGATGCTGCTGCGCTGCGCATACGAGGA-----TCTGTGCGC 1133
QY      1109  CTTGCTGCTGCGGAGAGCGGCGGACCTTCACCGGCTGCGCGCTCCGCGCGCGCGAGC 1168
Db      1134  ATTGCTTGTGCTGAGACCGCATGAGACAGTATCCATGCTTGTGTAACGGCTCCGCGC 1193
QY      1169  CGACCCCGCGCATGCTTTCGGAATCCCGAGCGCGCGCGCTTTCGCGCGCTTGCAGCC 1228
Db      1194  TTGCAATTCGCTGAGCGCTGCGGAGCATGAAATCGAGCTCAATCATGATGCTGTC-----1246
QY      1229  GCGCGCGCGCGACGACGAGAGCGCTTACTCTCAAGCGCATCGACGAGTGCCTCAAGAATCC 1288
Db      1247  -----GCGGCTACTTGAAGCAATCGAGCGCTTGAACCAACGCAATCG 1288
QY      1289  GCAACGCGGAGTGTGATGATGCTGCTGACCAACATGATGACCGCGCGCGAGCGGCA 1348
Db      1289  TCGATGCGGAGTGTGATGATGCTGCTGACCAACATGATGCTGCTGCTGCTGCTGCTG 1348
QY      1349  CGGCGCTGCGGCTTACTCTGCGCGCTGCGCGCATCAAGCCCGCTCCGCTGACGCGCGCTGC 1408
Db      1349  ATCATGATGATGCTTATCGCTACATGCGGAGGAGAACCCAGCGCGCTGCGGCGCTATT 1408
QY      1409  TCGAGTTCCTCCGAACTGTGCTGCTGAGCGCTTCCGCGGCGGCTTCTCAACATGCGCG 1468
Db      1409  TGGGTAACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1468
QY      1469  CCAGCGCGCGGCTGCTGATGCTCAAGCCCATCAAGGAGGAGGAGGAGGAGGAGGAGGAG 1528
Db      1469  GCCACGCGGAGATTCAGACCAAGCCCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1528
QY      1529  AGGAGAGCGAGCGGCTTCCGCGCGCTGCGCGCGCGCGGAGGAGGAGGAGGAGGAGGAG 1588
Db      1529  AACTGAGCGCTTAACTTGGCATGCGCTGCGCGCTGCGGAGGAGGAGGAGGAGGAGGAG 1588
QY      1589  TGTATGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1648
Db      1589  TGTATGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1648
QY      1649  ACGGCGCGCGGCTTCTGAGGCTGAGGAGCTTACGCGCGCGCGGAGGAGGAGGAGGAGGAG 1708
Db      1649  CCGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1708
QY      1709  TCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1768
Db      1709  TCCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1768
```

```
Db      1709  TGGAGCGAGGCTGCGCGCGCGGATTCGAGTCTGATGCACTGCTTAAAGCGGTCTCCCG 1768
QY      1769  GCGGCTTCATGACCGGCGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1828
Db      1769  GCGGCTTCATGACCGGCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1828
QY      1829  AAGGCGCGCGGCGCTTACTCTCGGCGCGCTCGAGTGTTCGCTTCAAGCGCGCGCGCG 1888
Db      1829  ATGCGCGCGGCTGCTGCTTATGCGGAGCATGCGCTACCTGCGCTTCAACTGCGCGCG 1888
QY      1889  ACTTCAGCATCTGATCCGACCATGCTGCTGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1948
Db      1889  ACTTAAACATTCGAGTCCGAGCTTCTTATGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1948
QY      1949  GCGGCGCGGATGCTGCTCTCTCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2008
Db      1949  GCGGCGCGGATGCTGCTCTCTCGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2008
QY      2009  CCGCGCGCATGCTCA 2023
Db      2009  CCGCGCGCATGCTCA 2023

RESULT 8
ADN36893
ID      ADN36893 standard; DNA; 55839 BP.
XX
AC      ADN36893;
XX
DT      15-JUL-2004 (first entry)
XX
DE      X. albilineas XALB1 gene cluster DNA.
XX
KW      Albicidin family; antibiotic production; biosynthetic gene cluster;
KW      XALB1; albicidin biosynthetic gene cluster 1; phytotoxic damage; gene;
KW      ds.
XX
OS      Xanthomonas albilineas.
XX
PN      WO2004035760-A2.
XX
PD      29-APR-2004.
XX
PF      17-OCT-2003; 2003WO-US031142.
XX
PR      18-OCT-2002; 2002US-0419463P.
XX
PA      (UYFL ) UNIV FLORIDA.
XX      (CIRA-) CIRAD CENT COOP INT EN RECH AGRONOMIQUE.
PI      Royer M, Gabriel DW, Frutos R, Rott P;
XX      WPI; 2004-365158/34.
XX
PT      New transformed host cell, useful for producing antibiotics, preferably
PT      polyketide antibiotics for protecting plants against phytotoxic damage,
PT      or damage against albicidin.
XX
PS      Claim 8; SEQ ID NO 1; 193bp; English.
XX
CC      The present invention relates to a novel Albicidin family of antibiotics
CC      produced by the expression of biosynthetic gene clusters from Xanthomonas
CC      albilineas designated as XALB1, XALB2 and XALB3 (albicidin biosynthetic
CC      gene clusters 1, 2 and 3). The invention discloses the polynucleotide
CC      sequences of these gene clusters, and the proteins encoded by the open
CC      reading frames (ORFs) within the gene clusters. Also disclosed are
CC      methods for producing an antibiotic and protecting a plant against damage
CC      from albicidin and against phytotoxic damage. The present sequence
CC      represents Xanthomonas albilineas XALB1 gene cluster.
XX
SQ      Sequence 55839 BP; 10970 A; 16145 C; 17705 G; 11019 T; 0 U; 0 Other;
Query Match      17.6%; Score 361.8; DB 12; Length 55839;
```


Best Local Similarity 51.4%; Pred. No. 4.8e-38;
Matches 1057; Conservative 0; Mismatches 932; Indels 66; Gaps 7;
QY 2 TGCGCAGCTTCTGATGCAACAACGACTGCTTACCCACACACTGTTCCAGTACATCG 61
Db 14910 TGGCTGCTCTTATCATMAATATTAATTCATGTTGCTGATCTGGAAATCTGGCAGCTACGAG 14969
QY 62 GCGAGGCCACCGGCAACCCCGCTGCTGCTGCCCAAGCAGCCGACTGTGCG----- 116
Db 14970 CGGAGATCTTCGGGAGAGATCCCTGGTGGGCAACAGAGAGATATCTCTGGACAGAAC 15029
QY 117 -GCTGCCCTTGAGAGACTTCAGCGGATGCTGTGTCCCGGGCCCCGGCAGCCGACC 175
Db 15030 TGAAGGACCGCGGGGATTTTCTCTGATCATCTTTGCGCCGGCTCCGGTCTGAGTTA 15089
QY 176 GGGAGCGGGACTTGGAATCAGCGCGGGGATCACCAGAGGGGCTGCGCTCTCG 235
Db 15090 ATGAAGGATTTTCAATCTCGCTGAGCGGCTGAGCAGAGAAATTTCCGGTGTAG 15149
QY 236 GCGTCTGCTCGGCAACGAGGCAATCGCCAGCTCTTCGGGCAACCGTCCGCTCGCC 295
Db 15150 GCGATGCTCGGGCTTTCAGGGAATTCGCAATGCTATGTTGGCCGATCTGATGCGC 15209
QY 296 CCGAACCATGACAGCGCCGGGCTTCGAGGTCGCGCACACCGGGAGAGAGTCTTCGGG 355
Db 15210 CCGTGCCCTTTCATGCGCGCTGCTCAACCGTCATCAACCGGAGCGGTTGTTGAG 15269
QY 356 GCGTCCCTCGCCCTTACCGCCGCTGCTACCACTCCCTGCGCGCAACGA-----CC 409
Db 15270 GCAATCCGAGCGCTTTCAGGCAAGTGGCTATCATCTGTTGATGCTTGGCAGCAATCGC 15329
QY 410 TCCCGCAGAGCTCGAACCCCTGCTGAGCGCAGCAGGGGCTGCTATGGGCTGCGGC 469
Db 15330 TGGCGCTGCTGTAAGTACAGCGCGGCTTACCAATTCGCTGTGTATGGGCTTGGAGC 15389
QY 470 ACCGCGAAGACCGCTGTGGGCGTCCAGTTCCACCCGAGTCCATCGGAGCACTTCG 529
Db 15390 ACGTGCAACACCGAAATGGGAGTACAGTTCCACCCGAAATGATCTTCACGAAACAG 15449
QY 530 GCGCGGAGATCATGAGCCACTTCGCGCACTCGCCCTCGGCCACACCGGCGACGCGCGC 589
Db 15450 GCAAGCGCATTTGTCTTACCTTTGCAAGCTGTGCTGCGCCACAGTGCACCGTTACTTG 15509
QY 590 ACGGGGCGAGCTCCCGCTAGCAACTCCAGTGGCGCGCTGAGCTGCTCCCGAGCGCG 649
Db 15510 CCGGGTGGAGCAGAGCGCGAAGTTTAAAGCTTTGCGCGCCGAAATGTATCAACCGC 15569
QY 650 AAGAGTACCGCGCGCTGCTGCGCCGCGAGGAGCACAGTTCTGAGTGA---CAGCA 706
Db 15570 GGGTACGTGCTGAGAGCGGAAAGATCAAGTGCCTTGGCAGGCGGAAGATGTCTTTC 15629
QY 707 GCTTCGCTCTGAAGCGCTCGCGCTTCTCTTCTCGGCGAC-----GACCGCGCGC 760
Db 15630 TGGCTTGTTCGCTGAGCAAAACATTTCTTGGCTGACAGCAGCTGATCTGAGATC 15689
QY 761 CGGTCCCGAGTACCTCACTACCGGTGCGCCAGCGGCTGTCTCGTTCGCGGCTCGC 820
Db 15690 CAATGGCGCGCTATTGCTTATGAGAGCGGTGAAACGAGACGAGGTATGTCCGATTTGCG 15749
QY 821 ACGGACACAGACCCCGGACGCGCGCCCTTCTTCAACTACCTGAGAGA----- 869
Db 15750 TGGGGCAGGAGACATGTGTGCAAGAGGCAAGCGAGGTTTCTTTCGTAAGATGATCGG 15809
QY 870 -GCAAGTCGAACCGCAGCGGCTCCCGTGGCCCCCGAACTGCTTCGAGTTCACTCG 928
Db 15810 CGTTGCAATCGGTCTTACTGAGAGAGTCCGCGAGCGGCAACCGTTTCGCGTTCGCGGG 15869
QY 929 GCTACGTCGGCTACTCTCGGTTAAGAGTGAAGCGAGACCAACCGCGAGACCCCGCGACC 988
Db 15870 GCTACGTCGGCTACTCAAGTGAAGTGAATCGGTTCGCGCGCGCTTCACATG 15929
QY 989 GGTCCCGCAACCCGAGCGCGGTTCTCTTTCGCGCAGCGGCGATTCGCTCGACACG 1048

Db 15930 CCAATGCCATCCCCGATGCGTTGAGATGCGCGGTGAGAGCGCTTCTGTTCCGACACG 15989
QY 1049 AGGAAGGCTGCTGTAACCTGCTGGCCCTCGAACCGCGGGGCGACAGAGCGGCGCGCG 1108
Db 15990 CCACTGAGAGTATGTTGTGCTGCGCTGCGCCATACGAGAGA-----TCTGTGGC 16041
QY 1109 CTTGCTGCGGAGAGAGCGGCGAGACCTTACCGGCTGAGCGCTGCGCGCCCGCGAGC 1168
Db 16042 ATTGGCTTGGCTAGACCCCATTCAGACAGCTATTCATTCGATTTGTCAACGGCTCGGC 16101
QY 1169 GCAACCCCGCATGTGTTCTTGGGATCCCGAGCGCGGCGGCTTCGCGCCCTGCGCC 1228
Db 16102 TTGCATTTTCGCTAGGCTCGCGCAGATGAAATCGAGCTCAATCATGTGTC----- 16154
QY 1229 GCGCGGCCACAGACAGAGAGCTTACTTCAAGCGCATTCAGAGTGTCTTCAAGAGATCC 1288
Db 16155 -----GCGGCTACTTGAAGCAATTCAGCGTTGCMAACACGATCG 16196
QY 1289 GCAACGCGAGTGTACAGAGATCTGCTGAAGCAATGTCACCGGCGCGAGAGCGA 1348
Db 16197 TCGATGGCGAGTCTTATGAATCTGTCTTACCACTGTTCTGTCTTCAGGCGGAGCTGG 16256
QY 1349 CCGCCCTGCGCTTACTCTCGCGCTGCGCGGCTCATCAGCCCGCTCCGTAAGCGCGCTGC 1408
Db 16257 ATCCATTTGATGCTCTATGCTACATGCGGAGAGAGAACCGAGCGGCTTCGGGGCTATT 16316
QY 1409 TCGAGTTCCCGGAATGCTGCTGCTGAGCGGCTGCGCCGAGCGGTTCTCATATGCGG 1468
Db 16317 TGGTAAACGTTACGATGATGATCTTATGTAATTCACACAGAGCGTTTCTGGAATGGAGC 16376
QY 1469 CGAGCGGCGCGCTGAGTCCAGAGCCCATCAAGGGAGACCGCGCCCGGGGGGAGACCGCGG 1528
Db 16377 GCCAGCGCAGATTCAACACCAACCAATCAAGGAGCACTGCGCGCTGCGAGAGATCCC 16436
QY 1529 AGGAGACGAGCGGCTCCGCGCGCATGCGGCTGCGGCGGAGAGAGACCGGCGGAGAAC 1588
Db 16437 AACTGACCGTAACTTGGCATGCGCGCTGCGCGCTTGGAAAAAGACCGAGCGGAAACT 16496
QY 1589 TGAATGCTCGAATGCTGCTGCGCAATGACCTTCAACAGCGTCTGCGGATCGCTCGTCC 1648
Db 16497 TGAATGCTCGAATGCTGCTGCGCAATGACCTTCAACAGCGTCTGCGGATCGCTCGTCA 16556
QY 1649 ACGTCCCGCGGCTCTGAGAGTGAAGCTTACCGCGCGCTGCAACAGCTGTGTCAGCA 1708
Db 16557 CCGTCCCAAGCTGATGACATCGAAAGCTTCAAGACCGTGCATTCAGATGTCAAGACCG 16616
QY 1709 TCGGAGGAGCGGCTGCGGCGCGGACACGACCGCGCTGCGTACCGCGCGCTTCCCG 1768
Db 16617 TGAAGCGAGGCTGCGCGCGCGATGCTGATGTGCACTGCTTAAAGCGGTCTCCCG 16676
QY 1769 GCGGCTTCATGACCGGCGCGCCCAAGAGCGCAATGAGATCATGACCGCTGAGAG 1828
Db 16677 GCGGCTCGATCAACCGCGCACCGAAGTTGCGAGTATGAGATTAATGATGCGCTGAGAG 16736
QY 1829 AAGGCCCCCGGCGGCTTACTCGGAGCGCTGAGATGTTTCGCTTCAAGCGCGCGCG 1888
Db 16737 ATGCGCGCGGTGCGGTATTCGCGAGCATCGGCTTACCTGAGCTTACAACTGCGTCCGCG 16796
QY 1889 ACCTGAGATGATGATCGGACACATGATGCTGCGGAGCGGCGGAGTTTCGCGTGC 1948
Db 16797 ACTTAAACATTTGCGATCCGCACTTTCTTATGACGGGAGAAATACGTTTCGCGCGCG 16856
QY 1949 GCGGGCGATGCTGCTCTCTCCGACAGAGAGAGATTCAACGAGACCTGTGTAAAG 2008
Db 16857 GCGGCGCATCAACCTTCTGTCGACACCGCAGAGATGAGTTTCAGACGAATGCTGTGAAG 16916
QY 2009 CCGGCGCATGTCTCA 2023
Db 16917 CCGAGCGATCTCTCA 16931

RESULT 9
AAS59535

ID AAS59535 standard; DNA; 26309 BP.
XX AAS59535;
AC
XX
DT 13-FEB-2002 (first entry)
XX
DE Propionibacterium acnes immunogenic protein encoding DNA #30.
XX
KM SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KM uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KM inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KM dermatological; osteopathic; neuroprotectant; ds.
XX
OS Propionibacterium acnes.
PN M0200181581-A2.
PD 01-NOV-2001.
PF 20-APR-2001; 2001MO-US012865.
PR 21-APR-2000; 2000US-0199047P.
PR 02-JUN-2000; 2000US-0208841P.
PR 07-JUL-2000; 2000US-0216747P.
XX
XX (CORI-) CORIXA CORP.
PI Sheiky YW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
DR WPI; 2001-616774/71.
PT Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX Claim 1; SEQ ID NO 30; 1069bp; English.
XX
CC Sequences AAS59506-AAS59804 represent DNA molecules encoding
CC Propionibacterium acnes immunogenic polypeptides. The proteins and their
CC associated DNA sequences are used in the treatment, prevention and
CC diagnosis of medical conditions caused by P. acnes. The disorders include
CC SAPHO syndrome (synovitis, acne, pustulosis, hypertosis and
CC osteomyelitis), uveitis and endophthalmitis. P. acnes is also involved in
CC infections of bone, joints and the central nervous system, however it is
CC particularly involved in the inflammatory lesions associated with acne
CC vulgaris. A method for detecting the presence or absence of P. acnes in a
CC patient comprises contacting a sample with a binding agent that binds to
CC the proteins of the invention and determining the amount of bound protein
CC in the sample. The polypeptides may be used as antigens in the production
CC of antibodies specific for P. acnes proteins. These antibodies can be
CC used to downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). This sequence encodes the
CC polypeptide shown in A4U6704-A4U6985 and A4U67509. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;
Query Match 15.3%; Score 316; DB 4; Length 26309;
Best Local Similarity 50.6%; Pred. No. 3.7e-32;
Matches 1057; Conservative 0; Mismatches 955; Indels 78; Gaps 9;
QY 10 CTTCTGATCGACACTGACCTGTTACCCACCACTGTTCCAGTACATCGGCGAGCC 69
DB CTGCTGTTGACACACCGATTCTTCACTTCACTTGGACACCTATTGGCCGAGATC 19074
QY 70 ACCGGGACACCCCGCTGCTGTCGCCAACGACGCCGACTGATCG----- 114
DB 19075 AGCGGATCGAGCGAGTGGTCCGCCCGGAGAGGTGAGGCTGACGAGGATCTCCGC 19134

QY 115 CGGCTGCCGTCGAGAGACTTCGACGCGATGTCGTGTCCCGAGCCCGACGCCGAC 174
DB CGATTGGCTGATGGGGAATTCCGACCGATGTTATGATCGACGACTTGGCTCCCGAC 19194
QY 175 CGGGAACGGGACTTGGGAATCAGCCGCCGGGCGATCA---CGACAGCGGCTGCCCC 231
DB 19195 AACGACAGGACTTCAAGACCCCGGACAGGTGATGATGCTCCAAAGCTCCCCCTG 19254
QY 232 CTCGGGTCGCTGGCCCGACGAGGCAATGCGCAGCTTGGGGGAAACGTCGCGC 291
DB 19255 CTCGGGGTGTGCTGGGCCACCAAGGCTCG--GGTTACGTACGGTCCCAATTGACG 19311
QY 292 GCCCGGAACCCATGACCGGCGGCTCCGAGTGGCGGACACCGCGGAGACGCTTTC 351
DB 19312 ACCATTGCGGCCCATATCAGGCGCATCGTCAGCCGAATTCATCTCCAGGGGTGAATCTTC 19371
QY 352 CGGGGCTTCCTCCGCTTCAACCGCGTGCCTACACTCCCTGCGCGCACCGACCTC 411
DB 19372 GCCGATATACCGCAAGATTTCGAGGCAACTCGGTATCACTGTCGTTGACACGTTG 19431
QY 412 CCGAGAGACTGGAACCCCTCGCTGAGAGCAAGACGGGGTGTATGGGCTTGGCGAC 471
DB 19432 GGTGACAGATCGTCGAACATCGTCGCGCGGAGGATGGCCATATGGCGTTGAGGTT 19491
QY 472 CGGAGAGCGCGGTGCGGCGCTCCAGTTCCACCCGAGTCCATCGGCGAGCACTTCGCG 531
DB 19492 GCCGACCGGCGACATCGGGGCGTGTGATTCACCCGAGATCACTGATGACGAGTCCGA 19551
QY 532 CGGAGATCATGGCCAACTTCGCGACCTCGCTTCGCCACACCGGGGACGCGCGAC 591
DB 19552 TGTCAATTATGACGAATTTCTTGGCGATCGGACGACGACGACACACGAGGTACCG 19611
QY 592 GGGGCGGACTCCCGCTGACGACTCC-----AGTGGCGCGCTGACGTGTCGCG 643
DB 19612 GACMAACGGGAACCGGCAACATCCAGTACGAAATGTGTCGCGGACTCGGCGAGTGG 19671
QY 644 AGCGGAGAGGTACGCGCGCG---CTGCTGCGCGGAGGACCAAGTTCTGGCTG 699
DB 19672 ACATATCACACCAACGATGTGACATCTGACGAAAGAACCACTTCAGCGGCTTC 19731
QY 700 GACGAGCTCGCTCTCTGCAAGCGCGCTTCCTTCCTTCGCGGACGACGCGGCG 759
DB 19732 GCTGGGAGCGGTGACGCTTGTGTGACTCGGCGACACGAGTGGCGACACCGGCGGT 19791
QY 760 CGGCTGCGGAGTACTCTACCTACCGCGTTCGCGGAGGCGGTGTCTCGCGGCTGC 819
DB 19792 TGGAGTGTATGGGCAAGCGGCTCGGGGCGGCAATCGAATGATGATGACGATCATC 19851
QY 820 GACGCGACACGACCGGGAACCGGCGCGCTTCCTTCACTTACCTGAGGAGGAGCTGGA 879
DB 19852 ACAACACGTTGACGGTCAACGGGCGAACCAGTCTGTGATGTGCTGGAATCTTTGAA 19911
QY 880 CGCGGACGGGTCCCGCTGCGCCCGCAACTGACCTTTCGA-----GTTCAAC 924
DB 19912 CGCGGCTTGTGAGATGAGATATACGAGAGGAGACCGGACATGATGATCCATTCAACA 19971
QY 925 CTCGCTACGTCGGCTAATCTCGGCTACGAGCTGAGGCGGAGACCAACGCGGACCCCGG 984
DB 19972 GGTGCTATGTGCGTTTCTTGCGGCTATGAGTCAAGCGCTGACCTTGGGCGGAAATGG 20031
QY 985 GACCGGTCCCGGACCGCGGACCGCGGTCTCTTCGCGGACCGCGGCAATCGGCTTCGAC 1044
DB 20032 CACTGAGGAAATGCGCGATGCAATGATGATGCTCGGCGACGCTGGAATGGTATCGAC 20091
QY 1045 CACGAGGAAGGCTGTGCTAATCTGCTGCGCTTCGACCGCGGGGCGACACGACGCGCGC 1104
DB 20092 CACTACCGGATCAAGCTCACTCTAGCTTGGAGAGACGCTGACACGCGCGCTGCAAT 20151
QY 1105 CGGCTGCGTGGGAGAGACGCGCGGACCTTCACGCGCTGAGCGTTCGCGCGCGCGCC 1164
DB 20152 GAGGCGCGTATC-----TATGAGACGACACTGCTGCGACGCTGCT 20192

OY	1165	GAGCCGACCCCCGCATGATCTTCTGGAGATCCCGAGGGGGGGCGGCTTCGGCCCCCTG	1224
Db	20193	GAGCGACCAAGTATACACCAAGAGCCCGACCGCTAACCGTCTCTGGAAAGGTCGTG	20257
OY	1225	GCCGCGCGCGCCACAAGACAGCGCCTTACTCTCAAGCGCATCGACAGAGTGCCTCAAGAG	1284
Db	20253	-----GGGGCTCAGCCGAGGGGGGCTACAGGACCGTGTTCCTCGGAATCCAGCAGGG	20306
OY	1285	ATCGGCACGGCGAGTGTATACAGATCTGCTTGAACCAATGTGTCAACGGCGCCGACCTGAG	1344
Db	20305	CTAACCGCGCGGTGACTCGTACAGAGCGCTGTCTGACCGACAGTGTGACCTCCGGTGCAG	20364
OY	1345	GCGACGGCGCTGGCGGCTCTACTCCGGCGCTGGCGGCCCATCAGCCCCCGCTACCGCGCC	1404
Db	20365	GTGACGGCTGGCGGCTGTATACCGGCACTGGGCGTGTGAACACCGCCCTCTATGAGGGCG	20422
OY	1405	CTGCTCGAGTTCCTCCGAATG---TCCGTGCTGAGCGCTCGCCCGAGCGGTTCTCTACG	1461
Db	20425	TATCTCGGTTTACCGACCGCGCTGTGAGGTGTGTCTTTCATCCCGGAACGGTTCGTG	20488
OY	1462	ATCGAGCGCGACGGCGGCGTGTGAGTCCAAAGCCATCAAGAGGACCGGCCCCCGGGCGCG	1521
Db	20485	CGAGTATCGCGACGGCATGTGCTGAATCCAGCCGATCAAGGGGACAATAGGTCGCTGTGAT	20544
OY	1522	ACCCGCGAGGAGAGACAGACGGCGCTCCGGCGGCACTGTGCGCGCGGAGGAAGACCGGGCC	1581
Db	20545	GATCCCGGTGGAGGATGCCCGGGGAATGTGTGATTTACACATGACCGGAAACCCCGCC	20604
OY	1582	GAGAACCTGATGATCTGTGTGACCTGTGTCCGCAAGCACTCAACACGCTGTGCGGATCGCG	1641
Db	20605	GAGAAATCTCATGATTCGTGATCTGTGTGCGACACACTGTGTGCGCGGTGTCCAACAGCG	20664
OY	1642	TCCGTCCACGTGCCCGCGCTCTTCGAGGTGGAGACTTACCGCGCGCTGTGACACAGCTGTGTG	1701
Db	20665	ACGGTGGAGGTACCAACGCTGTGATGGAGATGAGTCAATGCGACGGTGTGACACAGATGTGTG	20724
OY	1702	TCGACCATCGGGGACGGCTGTGCGGCCGACCAAGACCGCGCTGTGCGCGCGCC	1761
Db	20725	ACCAACGGTGCCTGGACCAATTGCTGAGAGAGCGTGGATTGTGTGACATCTGCGACGCAAC	20784
OY	1762	TTCCCGCGCGGCTCATGACCGGGCGGCGCAAGAGCGCAACATGAGATCATCGACCGCC	1821
Db	20785	TTCCCTGGTGGATGATGATGACCGGGGCAACGAAGGAAGGTCCGTGAAGAAATCTCGACGGT	20844
OY	1822	CTGAGAGAGGCGCCCGCGGCGTGTCTTACTCCGGGGCGCTCGAGATGTTCTGCCCTCAGCGCG	1881
Db	20845	CTGAGAGGTGCGCGCGCGGGGAATTTACTCGGGAATCTTGGGATATCTCGGCTTCAATCGC	20904
OY	1882	GCGCGCCGACTCAGCATGTGTATCTGCGACCAATGTGTGGCGCGACGGCGCGAGGTTC	1941
Db	20905	ACCCCGCAGCTCAACATCTGTGATCTGAGACGGTGTCCGAACGGGGTTCGACGGTAAACGGTTC	20964
OY	1942	GGCGTCCGGGGGGGATCTGTCTCCCTCTCCGACCAAGAGAGAGATTCAACCGAGACCGTG	2001
Db	20965	GGTCCCGCGCGGCAATCTGTGCAAGCTTTCACAGCTGTGACAGAGAGTGGCGCGAGAAAC	21024
OY	2002	GTAAAGCGCGCGCATGTGTACACCGGCTCGACGGGACGCGCGGTGGCGGG	2051
Db	21025	CTCAAAACGGGGGACCGCTTGGCTGGCTGACGCTGTCAAGCCGGGCAAG	21074
RESULT 10			
ACF64464			
ID	ACF64464 standard; DNA; 26309 BP.		
XX	ACF64464;		
XX			
DT	17-OCT-2003 (first entry)		
XX	Propionibacterium acnes DNA contig sequence #30.		
XX			
KM	Acne vulgaris; acne/seborrheic; dermatological; antibacterial; immunostimulant; immune response; vaccine; ds.		

XX	Proionibacterium acnes.
OS	
XX	WO200303515-A1.
PN	
XX	
PD	24-APR-2003.
XX	
XX	
PF	11-OCT-2002; 2002WO-US032727.
PR	
XX	
XX	15-OCT-2001; 2001US-00978825.
PA	(CORI-) CORIA CORP.
PI	Mitcham JI, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL,
PI	Zhang Y, Wang S, Yen S, Lodes MJ, Benson DR, Jones R, Carter D;
PI	Barth B, Valliève-Douglas J;
XX	
XX	WPI: 2003-381789/36.
DR	
PT	New Proionibacterium acnes polypeptides and polynucleotides encoding the
PT	polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT	or for stimulating an immune response specific for a P. acnes protein.
PS	Claim 1; SEQ ID NO 30; 1481bp; English.
XX	
CC	The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC	encoding a Proionibacterium acnes protein. The invention also relates to
CC	polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC	immunogenic fragments of P. acnes polypeptides. The invention
CC	additionally encompasses expression vectors and host cells comprising a
CC	polynucleotide of the invention; antibodies against polypeptides of the
CC	invention; fusion proteins comprising a polypeptide of the invention; a
CC	method for stimulating an immune response specific for a P. acnes
CC	polypeptide and an isolated T cell population comprising T cells prepared
CC	via this method; a vaccine composition (comprising P. acnes polypeptides,
CC	polynucleotides, antibodies, fusion proteins, T cell populations, or
CC	antigen-presenting cells that express the polypeptide); a method and kit
CC	for detecting or determining the presence or absence of P. acnes in a
CC	patient; and a method for inhibiting the development of P. acnes in a
CC	patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC	proteins, T cell populations or antigen-presenting cells that express the
CC	polypeptides are useful for diagnosing, preventing or treating acne
CC	vulgaris, or for stimulating an immune response specific for a P. acnes
CC	protein. The polynucleotides can also be used as probes or primers for
CC	nucleic acid hybridisation. The vaccine composition is useful for the
CC	stimulation of an immune response against P. acnes, or for treating acne,
CC	and the kit is useful for performing a diagnostic assay. The present
CC	sequence represents a P. acnes DNA contig which is specifically claimed
CC	in the invention. Note: The sequence data for this patent did not form
CC	part of the printed specification, but was obtained in electronic format
CC	directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX	
XX	Sequence 26309 BP; 5546 A; 8173 C; 7885 G; 4699 T; 0 U; 6 Other;
QO	
Query Match	15.3%; Score 316; DB 8; Length 26309;
Best Local Similarity	50.6%; Pred. No. 3.7e-32;
Matches 1057; Conservative	0; Mismatches 955; Indels 78; Gaps 9;
OY	10 CTTCTGATCGACAACCTAATCGACTGTTACCCACACCTGTTCCAGTAATCGCGAGGCC 69
DB	19015 CTGCTGGTGTGACAACCCAGATTCCTTCACTTCAACTGGACACCTAATTGGCCGAAGTC 19077
OY	70 ACCGGGGAACCCCCCGTCGTCGGGCCAACAGAGCGGCACTGTCG----- 114
DB	19075 AGCGGGATCGAGCCAGTGGTGGTCCGCCGGAAGAGTCAAGGTGCAGAGGATCTTCCG 19133
OY	115 CGGCTGCGCCGTCGAGACTTCGACGAGATCGTGTCTCCCGGGCCCCGGACCCGAC 174
DB	19135 CGATTGGCTGATGGCGAATTCGACACAGTGGTATCGGTTCAGAGACTTGGCTCCCGGCAC 19199
OY	175 CGGGAACGGGACTTTCGATCAAGCCCGGGGCGATCA---CCGACAGCGGCGTCCCGTC 231
DB	19195 AACGACAAAGACTTCAAGACCGCGGACAGAGTATGATGATGCTTCCAAAGATCTCCCTG 19254

QY	232	TTCCGAGCTCTGCTCCCTCGGCGCACAGGAGCATGCCAGTCTTCCGCGGAAACCGTCGGCTTC	291
Db	19255	CTCGGGGTGTGCTCGGGCGCAACAAAGGCGCTCG---GGTTACGTCAACGATGCCCAAGTTGCAG	19311
QY	292	GCCCCGGAACCAATGACACGGCGGGGATCTCCAGGTGGGCACACCGGCGAGGACGTCCTC	351
Db	19312	ACCAATTGGGCCCCCATCACGGCATGTGACGCCGAATTCATCTACTCCAGGCGGTGAATCTTC	19371
QY	352	CGGGGCGCTCCCGTCGCGGTTCAACCGCGGTGCGTACCACTCCCTGCGCGCAACCGACCTC	411
Db	19372	GCGGGATTACCGGAAGATTTGAGGCGAACTCGGATCACTGATGCTTGCACCAAGTTG	19431
QY	412	CCCGACGAGCTGCAACCCCTCGCTGGAAGGAGGACGGGGTGTGATCTGAGCGTCGCGGCAC	471
Db	19432	GGTAGCCAGATCGTCGAACATGCTCCGCGCGAGGATGGCCCTCATCGGTGTTCGAGGTT	19491
QY	472	CGCGGAAGAGCGGCTGTGGGGGCTCCAGTTCCACCGGAGTCATCGGACGACGACTTCGCGC	531
Db	19492	GCCGACGGGCGACACTGGGGGTGTGCAATTCACCGGAGTCAGTCAATGACGAGGTGGAA	19551
QY	532	CGGAGATCATGAGCAACTTCCGCGACCTCGCCCTCGCCACGACCGGGGACCGGCGCAC	591
Db	19552	TGTGAGTTAATGAGGAAGCTTCTGGCGATGGGAGCGCAGCAAGACAGCAGAGTACCC	19611
QY	592	GGGGCCGACTCCCGTACGAATCC-----ACGTGGCGCGGTGCAAGTCTGCGCG	643
Db	19612	GACAAACCGAACC GGCGCAATCAAGTCAAGAAAGTGTCCGGGACTACGCGCAGGTGG	19671
QY	644	ACGGCGAAGAGATAGCGCGCGG---CTGCTCGCGCGGAGGCGACCAAGTTCTGGCTG	699
Db	19672	ACAATCCACCAACAAAGGTGACATGACCTCGACAAAGATCCACCTTCAGCGCGCTC	19731
QY	700	GACAGCAGCTCCGTCCTCGAAAGCGCCTCGCGCTTCCTCTCTCGGCGACACCGCGCG	759
Db	19732	GCTGGCGACGCTGACGCTTCTGTGTTGAATCGGCGCACACGCGTGGCAACCGCGCT	19791
QY	760	CCGCTCGCGAGTACTCACTCACTACCGGCTGCGCGACCGCGCGTCTCCGTCGGCGCTCC	819
Db	19792	TGGAGTGTCAATGGGCGACGCGCGTGGGGGGGCGATCCGAACTGATTCAGATTCAGACGTATC	19851
QY	820	GACGGCACACGACCCCGGACGCGCGCGCCCTTCTTCAACTGCTGAGAGACAGCTCGAA	879
Db	19852	ACAAACACGTTGAGGATCAACGGGCGAAACCAAGCTGTGATGTGCTTGGAATCTTCTTGAA	19911
QY	880	CGCGCGAGGAGTCCCGTCGCGCGCGCGCGGAACTGCGCTCGA-----GTTCAAC	924
Db	19912	CGCGCTCTTGGAGATGAGATATCAACGAGGAGACCCGGAATTAATATCAATCCATTCA	19971
QY	925	CTCGGCTACGTCGGCTACTCTCGGCTACGAGCTGAAGCGGAGACCAACCGCGACCCCGCG	984
Db	19972	GGTGGCTATGTGCGGTTCTGTGGGCTTAAAGTGAAGCGCGCTGACCTTGGGCCCGGAATGG	20031
QY	985	CACCGGTCCTCCGCAACCCCGAGCGCGGTTCTCTTCCGCGACCGCGCATCGCCTCGAC	1044
Db	20032	CACTCAGCGAATCGCGCGATGCGCATGTGATGTGCTCGCGCAGCTGGAATTTGTGTACAG	20091
QY	1045	CACCAAGAGGCTGTCTGTACTCTGTGCGGCTCGACCGCGGGGCGACAGCAACCGCGCGC	1104
Db	20092	CACATCGCGCATCAAGCTCACTTCTTGAAGCTTGAAGCAACGTGACACGCCCGCTGCAAT	20151
QY	1105	CGCGCTTGGCTGCGGAGAGCGGCGGAACCTTCAACGAGCTTGCGCTCGCGCGCGCGCC	1164
Db	20152	GAGGCGCTGATC-----TATTGAGCGCACTGGCTGTGCAACGTGTGT	20192
QY	1165	GAGCGGACCCCGCCATGTCTTCCGAGTCCCGAGCGGCGCGGCTTCGCGCCCTCG	1224
Db	20193	GAAAGCAACAGTATGACAGCGAAGGCCCGACCGCTAACCGTGTCTGTGGAAGGATCGTG	20252
QY	1225	GCCCGCGCGCGCACGACAAAGACGCTTACTTAAGGCGATCGACGATGTCTTCAAGAG	1284
Db	20253	-----GCGGCTCAAGCGAGGGGCTTACAGAGCGTGTTCGCCAATCCAGCGAGGGG	20304

QY	1285	ATCCGCAACCGCCAGTCTGTACAGAGATCTGCTGACCAATAGTATCACCGCGCCGACCGAG	1344		
Db	20305	CTAACGCGCGGTGTACTGTACGAGGCGCTGTCTGACCGACCGTAGACTCTCCGCTGCGAC	20364		
QY	1345	GCAGCGGCCCTGCGCGCTCTACTCCGCGCTGCGCGGCATCAGCCCGCTCCGTACGGCGCC	1404		
Db	20365	GTGGAACGGCTGCGCGGTGTGTACCGCGCATCTGCGGCGTCGAAACCCACCCCTCATGGGGGG	20424		
QY	1405	CTGCTCGAGTTTCCCGGAATCG---TCGGTGTGAGGCGCTCGCCGACGCGTTTCTCAAG	1461		
Db	20425	TATCTTGGGTTTACCGCACCGCGCTGTGAGGAGTGTGCTCTTCAATCCCGGAAACGGTTCTG	20484		
QY	1482	ATCGGCGCCCGACGGCGGCGCTGAGTCTCAAGCCCATATAAGGGACCGCGCCCCGGGGCGGC	1521		
Db	20545	GATCCGCGTGGAGGATGCGCCGGGGAATTTGTGATTTTCAACATGACCGGAAAAACGGCGC	20604		
QY	1582	GAGAACCTGATGATCTGTGACCTGTGCTCCGACAGACCTTCAACAGCGTCTGCGCGATCGCG	1641		
Db	20605	GAGAACTCATGATCGCTGATCTGTGTGCGGACGACCTGTGCGGCGGTGTGCCAACAGAGC	20664		
QY	1642	TCGCTCCACGTGCCCGCGGCTCTTTCGAGGTGAGAACCTACGCGCCGTGACCAAGCTGTGTG	1701		
Db	20665	ACGGTTGAGGATACCACTCTGATGTGGCGATCGATCATATGCGACGGTGTGACCAAGATGTGTG	20724		
QY	1702	TCGACCAATCCGGGGGACGGCTGTGCGGCGCGGACAGACAGACCGCGCGCTGTACGCGCGCGC	1761		
Db	20725	ACCAACGGTGC CGGACGATTTCTGTGAGACGTGTGATTTGATGACGTACTGTGAGCGAC	20784		
QY	1762	TTCCCGCGCGCTTCATGACCGGCGCGCCCAAGAGCGACCAATGAGATCATCGACCGC	1821		
Db	20785	TTCCCTGTGGATGATGATGACCGGGGACCGAAGAAACGGTCCGTAGAAATCTTGAGCGGT	20844		
QY	1822	CTGGAAGAAAGCCCCCGGGCGCTGTACTCCGGGGCGCTCGATGTTGTCGCCCTCAGCGGC	1881		
Db	20845	CTGGAAGGTGCGCGCGCCGGGAAATCTACTCGGGAATCTGTGGATATCTCGGCTTTCATTCGC	20904		
QY	1882	GGCGGCCCACTCAAGCATCGTCAATCCGCAACCATGCTGTGGCGGACGGCCAGGCGGAGTTTC	1941		
Db	20905	ACCGCCCACTCAAGCATCGTGAATCGGACGGATATCCGAAACGGGGTGCAGGTTAAACGGTTC	20964		
QY	1942	GGCGTCCGCGGGGCGATCTGTCTCTCTCCGACCAAGAGAGAGATTCAACCGAGACCGTG	2001		
Db	20965	GGTCCCGGCGCGCAATCTGTGCAAGCTTCCACGCTGTGACGAGAGTGGCGGCAAGAAAC	21024		
QY	2002	GTAAAGCCCGCGCGCATGTGTACCGCCCTTCGACGCGACGCGCGGTGGCGGG	2051		
Db	21025	CTCAAAACGGGGGACCGTTGGCTGTGCGCTGACGCTGTCAACCGGGCGCAAG	21074		
RESULT 11					
ID	ACA42265	ACA42265 standard; DNA; 1362 BP.			
XX	ACA42265;				
XX	19-JUN-2003	(first entry)			
DE	Prokaryotic essential gene #23922.				
XX	Antisense; ds; prokaryotic essential gene; cell proliferation;				
KM	drug design; gene.				
XX	Pseudomonas aeruginosa.				
OS	WO200271783-A2.				
PN	03-OCT-2002.				
PD	21-MAR-2002; 2002MO-US009107.				
XX	PF				

XX	21-MAR-2001;	2001US-00815242.
PR	06-SEP-2001;	2001US-00948993.
PR	25-OCT-2001;	2001US-0342923P.
PR	08-FEB-2002;	2002US-00072851.
PR	06-MAR-2002;	2002US-0362659D.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Hasselback R, Ohlssen KL, Zyskind JW
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH,
XX
WPI, 2003-029926/02.
DR P-PSDB; ABU83395.

PT New antisense nucleic acids, useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs.

PS Claim 14; SEQ ID NO 30135; 1766pp; English

The invention relates to an isolated nucleic acid comprising any one of the 6213 antisense sequences given in the specification where expression of the nucleic acid inhibits proliferation of a cell. Also included are: (1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at http://wipo.int/pub/publ/published_pct_sequences

SQ Sequence 1362 BP; 243 A; 479 C; 430 G; 210 T; 0 U; 0 Other;

Query Match	13.1%	Score 270	DB 8	Length 1362
Best Local Similarity	58.7%	Pred. NO. 4	5e-26	
Matches 487	Conservative 0	Mismatches 340	Indels 3	Gaps 1

QY 1227 CCGCGGCGCCACGACAAAGGACCGCTTACTTCAAGCGATTCGACGAGTCCCTCAAGAGAT 1286
Db 528 CCGACGGAGCATCGCCCATCCGACTTATCGGCAAGCGCATCCGCGCATCCGAGCATCAT 587
QY 1287 CCGCAACGGGAGTCTGATCGAGATCTGCTGACCAACTGGTCAACCGGCGGACCGAGGC 1346
Db 588 CCAAGCGAGGAGACTGCTTACGAGGAACTATAGCCAGGCTTTCCAGGCGCTTGCAGCG 647
QY 1347 GAGGCGCCCTCCGCTTACTTCCGGGCTGCGGCGCATCAGCCCGCTCCGTTACGGCGCCT 1406
Db 648 CTGCGCCTGCGCGGCTTATCGCGGCTCGCGGAGGCTCGCCGACGCGGTTCTCCGGCTA 707
QY 1407 GCTCGAGTTCCCGCACTGTGCGTGAAGGCGCTCGCCGAGGCGTTCTTCAAGATCG 1466

Db	708	CCTGGACATGGCCGACCGCGCGGATCTCTAGCTCTGTGCGCGGAGCGCTTCTCAAGCTCGG	767
QY	1467	CGCCGACCGCGCGCTCGTAGTCAAGCCCATTAAGAGGACCTCGCCCTCCGGGCGGACATCCG	1526
Db	768	--CAAGGGCGAAGGTGGAAACCTCCGCGATCTAAGGGGACACCGCCCGCGCGCAAGACCC	824
QY	1527	GGAGAGGAGCAGAGGGGCTCCGCGCGCGACCTAGCCCGGCGGAGAAAGGACCGGGCGCGAA	1586
Db	825	CGAGAGGAGACATGGCGCTGGCGGGGTGCTGTGGCCAGGCCCCCAAGGACCGCGGGAAAA	884
QY	1587	CCTGATGATCGTCCGACCTGAGTCCGCAACGAACCTTCAMCAGCGTCTGCGCGATTCGCTTCGT	1646
Db	885	CCTATATATCGTGCACCTGCTGCGCGCAACGACATCGACGCGAGTTGGCCAACTGGCAGCGT	944
QY	1647	CCACGTGCCCCGCGCTCTTTCAGAGTGAAGACCTACGCGCCCGTGCACACGACTGTGTGCAC	1706
Db	945	ACGGGTACCGGAGCTGTTTGCCCTGGAAAGCTATCCCAACGTGCATCACTGTGTGAGCAG	1004
QY	1707	CATCCGGGGGACGGGTGGCGGCTCCGGCACACAGACACCGCGCTGCGTACGCGCCGCTTCCC	1766
Db	1005	CGTACACGGCGAACCTGGCGCGCGGCGCAAGGACGCGCTTGACCTGTGAGAAAGCAGCTTCCC	1064
QY	1767	CGCGCGCTCCATGACCGGCGCGGCCCAAGAAAGCGCACCATGAGATCAATGCACCGCCTTGA	1826
Db	1065	CGCGCGCTCGATCACCGGCGCGGCCGAAGATTTGCGCCATCAAGTCAATCGACGACTGGA	1124
QY	1827	GGAAGGCCCGCGGCGCTTACTTCCGGGGGCTCTCGATGAGTTTGCCCTCAGCGGCGCGC	1886
Db	1125	ACCGAGCGCAACCGCGGACCTTAAGTGGCGCAACCTGTCTTACTCGAGCGTGGCGGCGAGAT	1184
QY	1887	CGACCTCAGCATCGTCAATCCGACACATCGTGTGGCGGACCGGCCAGCGGAGTTCCGCGT	1946
Db	1185	GGACAGCTCGATCGCCATCCGCAACCTCGTGTGTGAGGAGCGGCGAGGTCAGTGTGTGGG	1244
QY	1947	CGCGGGGGCGATGCGTGTCCCTTCCCGACGACGAGAGGAGATTACCGGAGCCGCTGTAA	2006
Db	1245	CGGCGGGCGGATCGTCCGCGCACTGCGACTGGAAGAGACGATACGAGAAACCTCTGACAA	1304
QY	2007	GGCCCGCGGCATGATGTCACCGCCCTCGACGCGACAGCGCGTGGCGGGCGCC	2056
Db	1305	GGTCCGGGTGCTGCTGGAAACCTTGAGAGGAAATGGCCGGGACAGCGTCCC	1354

RESULT 12

ABD14809/c

ID ABD14809 standard; DNA; 1101 BP.

XX ABD14809;

XX 29-JUL-2004 (first entry)

XX

DE Pseudomonas aeruginosa polynucleotide #13413.

XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;

KW antibacterial.

XX

OS Pseudomonas aeruginosa.

XX

PN US6551795-B1.

XX

PD 22-APR-2003.

XX

PF 18-FEB-1999; 99US-00252991.

XX

PR 18-FEB-1998; 98US-0074788P.

XX

PR 27-JUL-1998; 98US-0094190P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;

XX

WP1; 2003-615309/58.

XX

P-PSDB; ABO81238.

XX

XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 13413; 455bp; English.
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-*P. aeruginosa* drugs, as templates for recombinant
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
CC infection, and in detection of *P. aeruginosa* sequences or other sequences
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-
CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1101 BP; 158 A; 355 C; 381 G; 207 T; 0 U; 0 Other;
Query Match 13.0%; Score 268.4; DB 11; Length 1101;
Best Local Similarity 58.6%; Pred. No. 7,7e-26;
Matches 486; Conservative 0; Mismatches 341; Indels 3; Gaps 1;
QY 1227 CCGCGCGCGCCAGACAGAGACCGCTACCTCAAGCCGATGACGAGTCTCAAGAGAT 1286
DB 862 CCGACGAGCATGACGCGCATCCACATTCGCGAGGCGATCCGCGATCCAGACTACAT 803
QY 1287 CCGCAAGCGGAGTCCGACAGATCTGCTGCACCAATATGTATCCGCGCCAGCCAGGC 1346
DB 802 CCAAGCAGCGACTGCTACCAAGAGATCTATAGCAGCGCTTCCAGGCGCTTGACAGCG 743
QY 1347 GACGCGCCCTCCGCTACTCCGCGCTGCGCGCATGAGCCCGTCCGTCAGCGCGCT 1406
DB 742 CTGCGCGTGGCCGCGCTATGCGCGCTGCGCGAGCGCTGCGCGCGCTTCTCCGCTA 683
QY 1407 GCTGAGTTCCTCCGAACTGCTGCTGAGCGCTGCGCGAGCGGTTCCTCAAGTCGG 1466
DB 682 CTGCGCATGGCGACGCGCGCATCTCAGCTGCTGCGCGAGGCTTCTCAAGCT--- 626
QY 1467 CGCGCAGCGCGCGCTGACGTCAGTCCAGCCCATCAAGGGGACCCGCGCGCGCGCACCC 1526
DB 625 CGACAAAGGCGCAAGGTGAAACCCGCGCGATCAAGGGGACCCGCGCGCGCAAGACCC 566
QY 1527 GGAAGAGACGAGCGGCTCCGCGCGCACTGCGCGCGCGGAGAGAGACCGGCGCGAGAA 1586
DB 565 CGAGAGAGACATGCGCTGCGCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGAA 506
QY 1587 CTTGATGATCGTCGACCTGCTGCGCAAGCACTCAACAGCGTCTGCGCGATCGCTCCGT 1646
DB 505 CTTGATGATCGTCGACCTGCTGCGCAAGCACTGCGCGCGCGCGCGCGCGCGCGCGCT 446
QY 1647 CCACGTGCGCGCGCTCTTCAAGTGGAGACCTACGCGCGCGCGCGCGCGCGCGCGCT 1706
DB 445 ACGGATACCGAGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 386
QY 1707 CATTCGCGGAGACGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCC 1766
DB 385 GGTACCGGCGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 326
QY 1767 CCGCGCGCTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1826
DB 325 CCGGCGCTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 266
QY 1827 GGAAGCGCGCGCGCGCTTACTCCGCGCGCGCTGCGAGTGTGCGCTTCAAGCGCGCG 1886
DB 265 ACCGAGCGCGAGCGCGCATCTTACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 206

QY 1887 CGACTCAGCATCTGATCCGACCATCTGCTGCGCGCGCGCGCGCGCGCGCGCGCT 1946
DB 205 GGACAGCTCGATCGCATCCGACCTGCTGCTGAGAAAGCGCGCGCGCGCGCGCGCG 146
QY 1947 CCGCGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2006
DB 145 CCGCGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 86
QY 2007 GCGCGCGCGCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2056
DB 85 GGTCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 36
RESULT 13
ABD15129
ID ABD15129 standard; DNA; 1431 BP.
XX
XX ABD15129;
AC
XX
XX 29-JUL-2004 (first entry)
DT
XX
DE *Pseudomonas aeruginosa* polynucleotide #13733.
XX
KW Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
KW antibacterial.
XX
OS *Pseudomonas aeruginosa*.
XX
PN US651795-B1.
PD
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
PI Rubenfield MJ, Nolling J, DeLoughery C, Bush D;
DR WPI; 2003-615309/58.
XX
XX P-PSDB; ABO81558.
PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 13733; 455bp; English.
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-*P. aeruginosa* drugs, as templates for recombinant
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
CC infection, and in detection of *P. aeruginosa* sequences or other sequences
CC of *Pseudomonas* species using biochip technology. Sequences ABD01397-
CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
CC seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1431 BP; 262 A; 503 C; 441 G; 225 T; 0 U; 0 Other;
Query Match 13.0%; Score 268.4; DB 11; Length 1431;
Best Local Similarity 58.6%; Pred. No. 7.3e-26;
Matches 486; Conservative 0; Mismatches 341; Indels 3; Gaps 1;

OY	1227	CCGCGCGCGCCACGACCAAGGAGCGCCCTACCTCAAGGGCATTCAGCGATGCTCCAAAGAGAT	1288
Db	597	CCGACGGAGCATCAGCGCATCCGACTTATCGGAGCGCATCCGCGCATCAGAGATTACAT	656
OY	1287	CCGCAACGGCGAGTGTGACAGATGTGCCTGACCAACATGTGTACCGGCGCGACCGAGGC	1348
Db	657	CCAGGACGACGCACTGCTACACAGGTGAATTATAGCCAGGCGTTCCAGGCGGCTTGCAGGCG	716
OY	1347	GACGGCCCTGCGGCTCTACTCCGCGCTGGCGGCATCAGCCCGTCCGTACGCGCCCT	1408
Db	717	CTCGCGGTGGCCGCGCTTATCCGCGCCCTGGCGAGGCGCTGCGGACGCGCTTCTCCGGCTA	776
OY	1407	GCTCAGTGTCCCGGAACGTCCGATGTGAGCGCCTCGCCGAGCGGTTCTCAAGATCG	1466
Db	777	CCTGGAGATTGGCCGACGCGCGGATCTTCACTGTGCGCGAGGCGCTTCTCAAGCT---	833
OY	1467	CGCCGACGGCGCGCGTGTAGTCCMAGCCATCAAGGGAGCCGCGCCCGCGGGCGAGCCGC	1528
Db	834	CGACACAGGGCAAGGGTGAACCCCGCGCATCAAGGGGACCCGCGCGCGGGAAGACCC	893
OY	1527	GGAGAGAGACGAGCGGCTCCGCGCCGACTTGCCGCGCGCGGGAGAAAGACCGGCGGAGAA	1586
Db	894	CGAGGAGGACATGGGGCTGGCGGCTCGCTGTGCGAGCCGCCAAGGACCGCGCGGAAAA	953
OY	1587	CCTGATGATCGTGCAGCTGTGTCGGGAAAGACCTCAACAGCGCTTTCGCGGATCCGCTCCGT	1646
Db	954	CCTGATGATTCGTGCAGCTGTGTCGGGAAAGACCTCAACAGCGCTTTCGCGGATCCGCTCCGT	1013
OY	1647	CCACGTGCCCCGGCTCTTTCGAGGTGAGACCTACGCGCCGCTGCACCACTGTGTGAC	1706
Db	1014	ACGGGTAACGGAGCTGTTGCCCCGTGGAAGCTATCCAACTGTGATCACTGTGTAGCAG	1073
OY	1707	CATCCGGGAGCGGCTGGCGCCCCGGGACACAGACACCGCGCTGTGCTGACGCGCGCTTCCC	1766
Db	1074	CGTCACTGGCGAATGGGGCGCGGGCAAGACGCGCTCGACTGTGGAAGGCAAGCTTCCC	1133
OY	1767	CGGCGGCTCCATGACCGGCGCGCCCAAGAGGACCAATGAGATCATGACCGCGTGA	1826
Db	1134	CGGCGGCTCGATCACCGGCGCGCGGAGATTTCGGCCATGACATCATGACGAATCGGA	1193
OY	1827	GGAAGAGCCCCCGGGCGCTGTACTCTCGGGGCGCTCGAGTGTTCGCTCTCAGCGCGCGC	1886
Db	1194	ACCGAGCGGACGCGGCGATCTACTGGGAGGCGCTGTTCTCACTCGACGTGGCGGAGAT	1253
OY	1887	CGACTCAGCATCGTCAATCCGACCATGTGTGCGCGACGCGCGGAGGTTTCGCGCT	1946
Db	1254	GGAACGTCTGATTCGCGATTCGCAACCTGCTGTGTGAGAAACGCGCGAGTCACTTCTGGGG	1313
OY	1947	CGGCGGGCGATCGTGTTCCTCTCTCCGACCAAGAGAGAGTTCAACGAGACGTTGTTAA	2006
Db	1314	CGGCGGGCGGATCTGTGCGCTGACTCGCATGTGGAGAGAGATGACAGAAACCTGTGACAA	1373
OY	2007	GGCCCGCGCATGTGTACCGGCTTCGACGAGGCGCGGTGGCGGGCGGCC	2056
Db	1374	GGTCCGGGTGCTGTGAAACCTCTGGAAGGATGACCGGAGCAGGATCCC	1423
RESULT 14			
ADB06077			
ID ADB06077 standard; DNA; 2046 BP.			
XX ADB06077;			
XX			
DT	20-NOV-2003	(first entry)	
XX			
DE	Alloicoccus otilis antigenic protein encoding DNA SEQ ID NO:17.		
XX			
KW	Alloicoccus otilidis; antigenic protein; immunogenic; immunisation;		
XX	gene therapy; Gram-positive bacterium; infection; gene; ds.		
XX			
XX	Alloicoccus otilis.		

Pt	WO2003048304-A2.
Pd	12-JUN-2003.
Pf	25-NOV-2002; 2002WO-US036123.
Pr	29-NOV-2001; 2001US-033777P.
Px	18-NOV-2002; 2002US-0426742P.
Pa	(AMHP) WYETH HOLDINGS CORP.
Pi	Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ;
Dl	WPI; 2003-505284/47.
Pt	P-PsDB; ADB06078.
Pt	New Alloiococcus otitidis polynucleotides and polypeptides, useful for
Pt	treating and diagnosing diseases, drug screening assays and monitoring of
Pt	effects during drug clinical trials.
Pt	Claim 7; SEQ ID NO 17; 1019pp; English.
Pt	The present invention describes an isolated polynucleotide (I) of
Pt	Alliotooccus otitidis genomic DNA, which encodes an antigenic protein.
Pt	Alliotooccus otitidis is a Gram-positive bacterium. Also described: (1)
Pt	an isolated polypeptide that is encoded by the polynucleotide (1); (2) an
Pt	expression vector comprising the novel isolated polynucleotide (1), its
Pt	complement, degenerate variant or fragment; (3) a genetically engineered
Pt	host cell, transfected, transformed or infected with the vector of (2);
Pt	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic
Pt	composition comprising the polypeptide, its complement, biological
Pt	equivalent or fragment, or the polynucleotide that is comprised in the
Pt	expression vector; (6) a pharmaceutical composition comprising the
Pt	polypeptide of (1) and a carrier; (7) a protein chip comprising an array
Pt	of the polypeptides of (1), their biological equivalent or fragment; (8)
Pt	immunising against Alliotooccus otitidis by administering to a host the
Pt	immunogenic composition; (9) detecting and/or identifying Alliotooccus
Pt	oitidis in the biological sample; (10) a kit comprising a container
Pt	containing the novel polynucleotide, its degenerate variant or fragment,
Pt	or the antibody of (4); and (11) producing a polypeptide by culturing the
Pt	genetically engineered host cell under conditions suitable to produce the
Pt	polypeptide from the culture. (1) can be used in gene therapy. The
Pt	polynucleotides, polypeptides, antibodies and compositions of the present
Pt	invention can be used for treating and diagnosing diseases, drug
Pt	screening assays and monitoring of effects during drug clinical trials.
Pt	The polynucleotides are useful for expressing and detecting Alliotooccus
Pt	oitidis. The present sequence encodes an Alliotooccus otitidis antigen
Pt	protein from the present invention.
SQ	Sequence 2046 BP; 597 A; 470 C; 493 G; 486 T; 0 U; 0 Other;
Query Match	13.0%; Score 267.2; DB 9; Length 2046;
Best Local Similarity	48.8%; Pred.No. 9.6e-26;
Matches 999;	Conservative 0; Mismatches 963; Indels 84; Gaps 7;
Oy	1 ATGGCGACGCTTCGATCGACAATACTAGACTCGTTACCCCAACCCTGTCCAATCATC 60
Db	22 ATGAAGTCACTCATTAATCGAATTAATGATTCTTAACTTAACAACTTAATACAGTTAAT 81
Oy	61 GGCGAGGCACACCGGCAACCCCCCGTGTCGCCCAACGACGCGCATGTCGGCGGTG 120
Db	82 GGAAGGTCAACGGAAGAAGAGCCATGATGATTAATAAAAAATGACACGATGACTTACCAAGA 141
Oy	121 CCCGTCA---GACCTTCGACGCGATCGTGTGTCGCCGGGCCCGGACGCCCAGACGG 177
Db	142 CTATTGATCTCGACTTGAATTAATGATCATTAATTTACCGGGGCCGGTAGTCCGAGACGG 201
Oy	178 GAACGGGATTCGGAATCACGCGCCGCGGCATCACGACGACGCGGCTGCTCTCGGC 237
Db	202 GACAAAGATTCGGCTGTGTGCGGCAAGTAATGAAGCTTGACAAAGCCATATTTTGGC 261
Oy	238 GTGTGCTCGGCCACGAGGATTCGACAGCTTCCTCGGCGGAACCGTTCGCGCTCGCCCG 297

Db 262 ATTGCTGGGCCACGAGGCATTATTATTATTAAGCGTGCACCTTAGTGGGGCGCTGAT 321
Qy 298 GAACCCATGACGCGCGGGGTCTCCGAGTGGCGCACCGCGGAGAGACGTCTTCGGGGC 357
Db 322 ATCCCATGATGGTGGCAGAGTGGCTTACCATTAATGGGAAAAATATCTTTGACGGC 381
Qy 358 CTCCCTCGGCTTACCGCGTGGCTACCACTCCCTGG--CCGCCACGACCTTCC 414
Db 382 ATAGACCAAGGCTTGAAGTCAACCGCTACCACTCCATTTGTGTGAAGACAAGAAATTA 441
Qy 415 GAGAGCTGACACCCCTCGCTGAGAGGACGAGGGGTGTCATGGGCTTCGGGCACCGC 474
Db 442 GAGATATTCAATTCAGTGCATAAAGCATGATGAATGTCAATGGCTGTCTCCACAG 501
Qy 475 GAGAGCGCTGTGGGGCGTCCAGTTCCACCCGAGTCCATCGGACAGCGACTTCGGCGG 534
Db 502 ACCAGCCATTCATGGGGTCCAATTCCACCCGAATCCATGGCACCGAGTATGGGGAA 561
Qy 535 GAGATCATGGCCAACTTCCGCGACCTGCGCTGCCACACACGGGGACGGCGCCACGGG 594
Db 562 AAATGATGAAAACTT-----TATGGCTTGTGCCAGGAC 597
Qy 595 GCGGACTCCCGTACGAATCCACGTGGCGCGGTGACGTGCTGCCGCGACGCGGAGAG 654
Db 598 TACTATAACCAATCCAGCTTTTATTATGAAAAAGTTCCGGAACTAGGGACAGCCAGAC 657
Qy 655 GTACGCGCGGCTGCTGCGCGGAGGACCAAGTTCTGGTGAACAGACTCCGCT 714
Db 658 CTCTAGCATCAATTGGCCAAATTTGATGACCAAGTACTTTGGCTCAGCTCAGCAAGTG 717
Qy 715 CTGAGAGGCGCTCGCGCTTCTCTTCTCGGCGACGACCGGCGCGCTCGCCGAGTAC 774
Db 718 GAGCTGGCTTGTCACGTTTCCATCTTGGCATGCTGGAACCAAGCGGGCCACAC 777
Qy 775 CTACCTAAGCGGCGCGGACGAGGTGCTCCGTCGCGGCTCCGACGCGCACACAGAC 834
Db 778 TTTAAATATCAGATGTCGACAAAAAGAAATGAAAAAAGATAGCCAGTCCGACAGCTT 837
Qy 835 CGGACGCGGCGCCCTTCTTCAACTAGAGGAGCAGCTCGAACGCGCAGCGGATCCC 894
Db 838 GAAA---GTTCCAGATGACATCTTTCTTAACTTAAAGCCAAACGACCAAAATGGCCT 894
Qy 895 GTGGCCCCGAACTGCGCTTGAAGTTCACTCGGCTACGTGGCTACCTCGGCTACGAG 954
Db 895 TATGAGCGGCTTGGCTTGAATTTCACTAGGCTATATGGCTATATCGCTATAGAG 954
Qy 955 CTGAAGCGGAGAC---CACCGGAGACCGCGGACCGGTCGCCGACCCCGACCGCGCG 1011
Db 955 GTCAAAAAAATACAGTTCAAGTCCACAACAAGACCACTATCCGATGCTTAC 1014
Qy 1012 TTCTCTTTCGCGACGCGCATCGCCCTCGACACACAGAGAGCTGCTACTGCTG 1071
Db 1015 TTTACTATTGTGACCGGGCTTGTGTATGACCAACAGAGAGACCTTTACTTCTC 1074
Qy 1072 GCGCTGACCGCGGGGCAACGACGAGCGCGCGCTGGCTGGCGGAGACGCGCGAG 1131
Db 1075 TCCATCCA-----GGAATGAC 1089
Qy 1132 ACCCTACCGGCTGGCGGTCCCGCGCGCGGACCGGACCCCGCATGATCTTCGGG 1191
Db 1090 CAGATTGATTCAAGGGTCAAAAGACAGCTCAAGAGGCACTCAAAAGTTCAAGCAAA 1149
Qy 1192 ATCCCGAGGCGGCGCGGCTTGGCGCCCTCGGCGCGCGCGCAACAGAGAGCGC 1251
Db 1150 GGACCAAGAGGCAAGACTTTCGACGACTGGCTTCG---TCAAGCAAAAGCAGCC 1206
Qy 1252 TACCTCAAGGCATCGACGAGTGTCTCAAGAGATCCGAAAGCGGAGTGTAGAGATC 1311
Db 1207 TATATCAAGGATGTAAGCAATCAGACCTGATCAAGGCTGGGAAAGTTATGAAGTT 1266
Qy 1312 TGCTGACCAATATGTCACCGCGCGACGAGGCGACGCGCTCGCGCTTACTCGCGG 1371
Db 1267 TGCTTGACCAACCGGCTGATATTGAGGTCAGATTGATGCGCTCGCTTATTACAAACAC 1326

Qy 1372 CTGCGGCCATGAGCCCGCTCCGTAAGGCGCGCTGCTGAGTTCCCGCAACTGCGGTG 1431
Db 1327 TTGCGCGAAGAAAGTCCCGGTAGTACTGCTTTTTCCTTTAGATGATTTTCAAGTT 1386
Qy 1432 CTGAGGCTCGCGCGAGGCTTCTCAGATCGCGCGGACGCGGCTCGAGTCAAG 1491
Db 1387 CTTTCTCTTTCATGAGGCTTTTATATGTCAGACAGAGACCGGCTTGTGACACTAAG 1446
Qy 1492 CCCATCAAGGGAGCCCGCCCGCGGCGGACACCGGAGAGAGACAGCGGCTCCGCGC 1551
Db 1447 CCCATAGGGAGACGCTCAGACGGGGACAGATGACCAAGAAACCAAGACTTGATAGA 1506
Qy 1552 GACCTGCGCGCGGAGAAAGACCGGCGGAACTTGATGATGCTGACCTGCTCGC 1611
Db 1507 GGCCTCGGAGTGAAGAAACCAAGGCGAGAAACTGATGATGTGACCTTGGCGC 1566
Qy 1612 AACGACTCAACAGCTCTGCGCATCGGCTCCGTCCACAGTCCCGGCTTTCAGAGTG 1671
Db 1567 AATGATCTGGGCTGTTTGTGAAATCGGGTGGTGAAGTACCAAACTGATGATGTA 1626
Qy 1672 GAGACTTACGCGCGGTGACACGAGTGTGTCACCATCGGGGACGCTGCGCGCGC 1731
Db 1627 GAGACTTACTTCACTCCACTCCACAGCTGTGTTACAGGCTTTCAGCGCGGATCAAGAGAC 1686
Qy 1732 ACCAGCACCGCGCTGCTGACGCGCGCTTCCCGGCGGCTTCATGACCGCGCGCC 1791
Db 1687 CTAGATGTGGTGAAGTGTAGAAAAACACTTCCCGGGCGGGTCACTGAGCTGGGGCCT 1746
Qy 1792 AAGAGCGCACCATGAGATCATCGACGCTCGAGAGAAAGCCCGGGGCTCTACTCC 1851
Db 1747 AAAAAAAGAACCTGGAATTTATGATGACTTGAAGAGTTCAGAGAGGATCTAATCT 1806
Qy 1852 GGGGCGCTCGAATGTTGCGCCTCAGCGGCGCGCGGACCTGACATGCTCATCGCAC 1911
Db 1807 GGGACTATTGGCTTTTATGACCAATTCACATATGACTTCAACATGCTATCCGAGCC 1866
Qy 1912 ATCGTCTGCGGACGCGGACGCGGAGTTCGCGTGGCGGGGAGATCGTGTCCCTGCC 1971
Db 1867 AGTGTGTGAAGCAGACAAAGGCAAGCATCGGCTGGGGGTGCGATGTGTATCT 1926
Qy 1972 GACCAAGAGAGAGATTCACCGAGACCGTGTGAAGCCCGCGCATGTCACCGCCCTC 2031
Db 1927 GATCTGAAGAAAGATTGATGAAATGTTTAAAAAGCTAAGGGGCGCTTGTCCGCTTG 1986
Qy 2032 GACGCG 2037
Db 1987 CAGGCC 1992

RESULT 15
ADBI2064.00/c
WP Sequence split into 18 fragments Locus ADBI2064 Accession Adbi2064
WP Fragment Name Begin End
WP ADBI2064_00 1 110000
WP ADBI2064_01 100001 210000
WP ADBI2064_02 200001 310000
WP ADBI2064_03 300001 410000
WP ADBI2064_04 400001 510000
WP ADBI2064_05 500001 610000
WP ADBI2064_06 600001 710000
WP ADBI2064_07 700001 810000
WP ADBI2064_08 800001 910000
WP ADBI2064_09 900001 1010000
WP ADBI2064_10 1000001 1110000
WP ADBI2064_11 1100001 1210000
WP ADBI2064_12 1200001 1310000
WP ADBI2064_13 1300001 1410000
WP ADBI2064_14 1400001 1510000
WP ADBI2064_15 1500001 1610000
WP ADBI2064_16 1600001 1710000
WP ADBI2064_17 1700001 1754382
ID ADBI2064 standard; DNA; 1754382 BP.

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 21:04:23 ; Search time 382.352 Seconds
(without alignments)
8820.059 Million cell updates/sec

Title: US-10-089-514-1
Perfect score: 2061
Sequence: 1 atgcgcgcgtctctgacga.....ccgtggcgcgcgcgcgcgtga 2061

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/ptodata/1/ina/5A COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PTUS COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	820	39.8	2220	3 US-08-765-907A-14	Sequence 14, Appl
2	820	39.8	2220	4 US-09-987-614A-14	Sequence 14, Appl
3	459	22.3	4496	3 US-08-765-907A-6	Sequence 6, Appl
4	459	22.3	4496	4 US-09-987-614A-6	Sequence 6, Appl
5	365.4	17.7	2888	3 US-08-765-907A-1	Sequence 1, Appl
6	365.4	17.7	2888	4 US-09-987-614A-1	Sequence 1, Appl
7	272.2	13.2	645	3 US-08-403-852D-9	Sequence 9, Appl
8	272.2	13.2	645	4 US-08-510-646B-9	Sequence 9, Appl
9	272.2	13.2	645	3 US-09-231-818-9	Sequence 9, Appl
10	272.2	13.2	645	4 US-09-635-359B-9	Sequence 9, Appl
11	268.4	13.0	1101	4 US-08-252-991A-13413	Sequence 13413, A
12	268.4	13.0	1431	4 US-09-252-991A-13943	Sequence 13943, A
13	260	12.6	816	4 US-09-252-991A-13943	Sequence 13943, A
14	232.6	11.3	1473	4 US-09-902-540-5760	Sequence 5760, Ap
15	232.6	11.3	54484	4 US-09-902-540-5760	Sequence 1272, Ap
16	226	11.0	1308	4 US-09-902-540-8203	Sequence 8203, Ap
17	226	11.0	7760	4 US-09-902-540-845	Sequence 845, App
18	197.8	9.6	1506	4 US-09-252-991A-811	Sequence 811, App
19	196.6	9.5	1518	4 US-09-489-039A-6879	Sequence 6879, Ap
20	182.4	8.9	4403765	3 US-09-103-840A-2	Sequence 2, Appl
21	182.4	8.9	4411529	3 US-09-103-840A-1	Sequence 1, Appl
22	161.2	7.8	672	4 US-09-252-991A-796	Sequence 796, App
23	154.2	7.5	1356	4 US-09-252-991A-13332	Sequence 13332, A
24	154.2	7.5	1614	4 US-09-252-991A-13021	Sequence 13021, A
25	154.2	7.3	2694	4 US-09-252-991A-12457	Sequence 12457, A
26	150.8	7.3	13613	3 US-09-105-537-3	Sequence 3, Appl
27	150	7.3	3065	3 US-09-199-637A-129	Sequence 129, App

28	147.6	7.2	44377	2 US-08-804-227C-7	Sequence 7, Appl
29	147.6	7.2	44377	2 US-08-804-198-1	Sequence 1, Appl
30	143.8	7.0	4411529	3 US-09-103-840A-1	Sequence 1, Appl
31	141.6	6.9	570	4 US-09-902-540-8201	Sequence 8201, Ap
32	140.4	6.8	627	4 US-09-252-991A-10862	Sequence 10862, A
33	140.4	6.8	1326	4 US-09-252-991A-10823	Sequence 10823, A
34	139.4	6.8	4403765	3 US-09-103-840A-2	Sequence 2, Appl
35	138.8	6.7	1584	4 US-09-489-039A-497	Sequence 497, App
36	133.8	6.5	77536	4 US-09-410-551B-1	Sequence 1, Appl
37	133.8	6.5	77536	4 US-09-940-316B-1	Sequence 1, Appl
38	132.8	6.4	462	4 US-09-252-991A-13836	Sequence 13836, A
39	131.4	6.4	8438	1 US-07-945-283-1	Sequence 1, Appl
40	128.4	6.2	2219	3 US-08-510-646B-17	Sequence 17, Appl
41	126.2	6.1	15872	3 US-09-105-537-1	Sequence 1, Appl
42	126.2	6.1	15872	4 US-09-091-609-1	Sequence 1, Appl
43	126.2	6.1	15872	4 US-09-091-609-3	Sequence 3, Appl
44	123.8	6.0	5392	2 US-08-403-852D-1	Sequence 1, Appl
45	123.8	6.0	5392	3 US-08-510-646B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-765-907A-14
; Sequence 14, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMES-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DITRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2220
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-14
Query Match 39.8%; Score 820; DB 3; Length 2220;
Best local Similarity 64.4%; Pred. No. 1.7e+126;
Matches 1356; Conservative 0; Mismatches 655; Indels 93; Gaps 5;
2 TGGCAGCGTTCTGATGACAACTAGACTCTTACCCACCAACTCTTCAATACATCG 61
71 TGGAAACCTCGCTGATGACAACTAGACTCTTACCAACACTCTTCAATACATCG 130
62 GCGAGCGACCGGCAACCCCGTGTGTCGCAACGCGGACCTGTCGCGCTGC 121
131 CCGAGTGACGCGCGCGCGCTCGCTGTCGCAACGACACCGACCTGCGACG 190
122 CCGTCGAG-----GACTTGACGCGCATGCTGTGTCGCGGCGCGGACCGGAC 175
191 CCTGCGCGCGCGCGCTGTCGCAACGTCGTGTCACCGGCGCGGACCGCGGCA 250
176 GGAACGGAATCTGGAATACGCGCGCGGACCTACCGACGCGGCTGCGCTCG 235
251 CCGACACCGACCTGCGCGCTGCGCGGCGGATGTCGCAATGCGACCTGCGCTCG 310
236 GCGTCGCGCGCGCGCGCGGCGGATGCGCGCTTTCGCGGAAACGTCGCGCTCG 295

Db 311 GGGTGTGCTGGGCGACAGGACCCCTGTGCTGCGCGGCGCGCCGCTGTCCACGAC 370
Qy 296 CGGAACCCATGACAGCGCGGGGATCTCGAGGTGCGGACACCGGGAGAGAGATCTTCGGG 355
Db 371 CGGAACCCATGACAGCGCGGGGATCTCGAGGTGCGGACACCGGGAGAGAGATCTTCGGG 430
Qy 356 GCGTCCCTGCGCGGTTACCGCGGTGCTTACCACTTCGCGCGGACCGACCTCCCG 415
Db 431 ACATCCCTCCCGCTGACCGTGGTCCGCTACCACTGCTGACCGTCCCGGCACTGCGC 490
Qy 416 AGGAGCTGAAACCCCTGCGCTGAGAGCGACGAGGGGTGCTATGAGGCTGCGGACCGC 475
Db 491 CGGACCTGCGCGCGACCGCGACCGCGAGCGGAGCTGATGCGCGTCCCGCGCGC 550
Qy 476 AGAAGCGCTGTGGGGGCTGCAAGTTCCAGCGGAGTCCATCGGAGGACCTTCGGCGGG 535
Db 551 ACCTGCGCCGCTTCCGCTGCTGCAAGTTCCAGCGGAGTCCATCGGAGGACCGCGCACC 610
Qy 536 AGATCATGAGCAACTTCGCGGACCT----- 560
Db 611 GGATGCTCGGCAACTTCGCGGACCTGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCA 670
Qy 561 -----CGCCCTCGCGCGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 604
Db 671 CGGAACCGATACCGCGACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 730
Qy 605 -----CGTACGAATCTCAAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 646
Db 731 CGTCCGCGCGGTTGAGGAGGAGTACCGGCTGCAATGTCGCGAGTGGCTGCGTCCGCGAG 790
Qy 647 CGGAAGAGTACCGCGCGGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 706
Db 791 CGGACGCGCGGTTACCGCGCGCTGTTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCG 850
Qy 707 GCTCCGCTCTGAGAGGCGCTGCGGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 766
Db 851 GCGCGCTGAGCGCGGCTGCGCGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 910
Qy 767 CCGAGTACTCTACCTACCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 826
Db 911 GCGAACGATCTACTTACGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 970
Qy 827 CCAAGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 886
Db 971 GCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1027
Qy 887 GGGTCCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
Db 1028 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1087
Qy 947 GCTACGAGCTGAGAGCGGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1006
Db 1088 GCTACGAGACCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1147
Qy 1007 CCGCGTTCCTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066
Db 1148 GCGCGCTTCAATGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1207
Qy 1067 TGTGCGCTCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1126
Db 1208 TCTTGAGACTGAGACGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1252
Qy 1127 CCGAGACCTTACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1186
Db 1253 CCGAAGCTTGGCTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1312
Qy 1187 TCGGAGATCCCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1246
Db 1313 TCAACCTGCTGCGCGAGACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1372
Qy 1247 ACCTGCTACTCAAGCGGATGAGAGTGGCTCAAGAGATCCCGAGCGCGGAGTCTGATG 1306
Db 1373 CCGGCTACCGGGAATGCTGAGGAGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1432

Qy 1307 AGATCTGCTGACCAATGATGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACT 1366
Db 1433 AGGTGCTCTGAGGAATGCTCTCGGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1492
Qy 1367 CCGCGCTGCGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1426
Db 1493 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1552
Qy 1427 CCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1486
Db 1553 CCGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1612
Qy 1487 CCAAGCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1546
Db 1613 CCAAGCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1672
Qy 1547 GCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1606
Db 1673 AGGCTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1732
Qy 1607 TCGGAGAGACCTGCAACAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1666
Db 1733 TCGGAGAGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1792
Qy 1667 AGGTGAGACTTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1726
Db 1793 AGGTGAGAGACTTAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1852
Qy 1727 CCGGACCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1786
Db 1853 CCGGACCGTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1912
Qy 1787 CCGCGGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1846
Db 1913 CCGCGGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1972
Qy 1847 ACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1906
Db 1973 ACTCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2032
Qy 1907 GCAACATCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1966
Db 2033 GCAACATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2092
Qy 1967 TCTCGGAGCGAGAGAGGATTCACCGAGACCGTGTAAAGCGCGCGCGCGCGCGCGCGCG 2026
Db 2093 TGTGCGAGCGCGAGCGAGGTCGCGGAATGCTCTCAAGGCGCGAGACCACTTCGCG 2152
Qy 2027 CCT 2030
Db 2153 CCT 2156

RESULT 2
US-09-987-614A-14
; Sequence 14, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMBAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSETT, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A

CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US/08/765,907
FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 2220
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-14

Query Match 39.8%; Score 820; DB 4; Length 2220;
Best Local Similarity 64.4%; Pred. No. 1,76-126;
Matches 1356; Conservative 0; Mismatches 655; Indels 93; Gaps 5;

QY 2 TGCGCAGGCTTCTGATGGAACAACGACTGCTTCCACCAACACCTGCTTCAGATACATCG 61
DB 71 TCGCAACCTGCTGCTATGACAACGACTGCTGTTACCTACACCTCTTCAGATGCTGG 130

QY 62 GCGAGGCCACCGGGCAACCCCGCTGCTGTCGCCAAGACGCGGACTGTCGGGCTGC 121
DB 131 CCGAGGTGAAGCGGGCGGCTCCGCTGCTGCTGCCAAGACGACACCCGACCTGGAGG 190

QY 122 CCGTCGAG-----GACTTGAAGCGGATGCTGTCCTCCGGGCGCCGAGCCCGAGCC 175
DB 191 CCGTGGCGCGGGGCGACTTCGACAAAGTCGTCGTCACCCGGCCCGGACCCCGGCA 250

QY 176 GGGAAACGGGACTTGGGAATCAGCCGCGGGGATCAGACAGGAGGCTGCGGCTCG 235
DB 251 CCGACACCGGACTGGGGCTCAGCGCGGGGTGATCAGCAATGGGAACCTGCGGCTGCTG 310

QY 236 GCGTCTGCTGCGGCAACGAGGCAATCGCCAGCTCTTCCGGGGAACCGTGGGCTCGCC 295
DB 311 GGGTGTGCTGAGGCAACGAGGCTGTCCTGCTGCGCGGGCGCGGCTGCTCAACGAC 370

QY 296 CCGAACCCTATGACAGGCGGGGCTTCGAGGTGCGGCAACCGGAGGAGACTCTTCCGGG 355
DB 371 CCGAACCCTTTCAGGCGCGGACGACGACATCGGCAAGAGGAGGAGGCTGTCGCA 430

QY 356 GCGTCCCTGCTGCGGCTTACCGGCGGTGCTTACCTCCCTGCGGCGGCAACGACTCCCGC 415
DB 431 ACATCCCTTCCCGCTACCGTGTGCTGCTGCTGCTGCGGCGGCTGCTGCGGCTGCTG 490

QY 416 ACAGCTCGAACCCTGCTGAGCGAGCGAGGCTGATGAGGCTGCGGAGCGG 475
DB 491 CCGAAGCTGCGGCGGCAACCGGCGGAGCGGCGGAGCTGATGAGGCTGCGGCGGCGG 550

QY 476 AGAAGCGGCTGAGGCGGCTGCAAGTTCACCGGAGTCACTGAGCGAGGACTTCCGGG 535
DB 551 ACGGCGCGGCTTGGGCTGAGTTCACCGGAGTCACTGAGCGAGGAGGAGGAGGAGG 610

QY 536 AGATATGCGCAACTTCCGCGACT----- 560
DB 611 GGAATGCTGCGCAACTTCCGCGACTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCA 670

QY 561 -----CGCCCTCGGCGGCAACCGGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCA 604
DB 671 CCGAAGCGATACCGGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 730

QY 605 -----CGTAAGACTTCCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 646
DB 731 CGTCCGCGCGGCGGAGGAGTACCGGCTGATGAGTGGCGGAGGCTGCTGCTGCGGCGG 790

QY 647 CCGAAGAGTACCGCGGCGGCTGCTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 706
DB 791 CCGAGCGCGGCTTACCGGCTGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 850

QY 707 GCTCCGCTCTCGAAGGCGGCTGCTGCTTCTTCTGCGGAGGAGGAGGAGGAGGAGGAGG 766
DB 851 GCGCGCTGAGCGGCGGCTGCGGCTTCACTTCTGCGGCGGCGGCGGCGGCGGCGGCTG 910

QY 767 CCGAGTACTTACCGGCGGCTGCGGAGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 826
DB 767 CCGAGTACTTACCGGCGGCTGCGGAGGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGG 826

DB 911 GCGAAGATCACTTACGAGTCCGACCGGCGGCTGCGTCAAGAGAGGCTTACGCG 970
QY 827 CCAAGACCGGAGCGGCGGCGGCTTCTTCACTTCACTGAGGAGGAGGAGGAGGAGGAGG 886
DB 971 GCGAGACCGGCGGCGGCGGCGGCTTCTTCACTTCACTGAGGAGGAGGAGGAGGAGGAGG 1027
QY 887 GGGTCCCGTCCGCGGCGGAGTCCGCTTCACTTCACTGAGGAGGAGGAGGAGGAGGAGG 946
DB 1028 GCGGCGGCGGCGGCGGCGGCTGCTTCACTTCACTGAGGAGGAGGAGGAGGAGGAGGAGG 1087

QY 947 GCTACGAGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1006
DB 1088 GCTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1147

QY 1007 CCGGCTTCTTCCGCGGCGGCGGAGTCCGCTTCACTTCACTGAGGAGGAGGAGGAGGAGG 1066
DB 1148 GCGCTTCACTTCCGCGGCGGAGTCCGCTTCACTTCACTGAGGAGGAGGAGGAGGAGG 1207

QY 1067 TGTGAGCCTTCAAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1126
DB 1208 TCTTGGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1252

QY 1127 CCGAGACCTTCAAGCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1186
DB 1253 CCGAAGCTGAGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1312

QY 1187 TCGGAGTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1246
DB 1313 TCACCTTGTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1372

QY 1247 ACGCTTACCTCAAGCGGCTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1306
DB 1373 CCGCTTACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1432

QY 1307 AGATGCTGAGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1366
DB 1433 AGGTGCTGAGCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1492

QY 1367 CCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1426
DB 1493 GCGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1552

QY 1427 CCGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1486
DB 1553 CCGTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1612

QY 1487 CCAAGCCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1546
DB 1613 CCAAGCCATCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1672

QY 1547 GCGCGAGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1606
DB 1673 AGGCTCTCTCGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1732

QY 1607 TCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1666
DB 1733 TCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1792

QY 1667 AGGTGAGAGCTTACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1726
DB 1793 AGGTGAGAGCTTACGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1852

QY 1727 CCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1786
DB 1853 CCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1912

QY 1787 CCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1846
DB 1913 CCGGCAAGCACTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1972

QY 1847 ACTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1906
DB 1973 ACTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2032

```
QY 1907 GCAACATGCTGCTGCGCCGACGCGGAGTTGCGCGTCCGCGGCGCATGTCCTCC 1966
      |||||
Db 2033 GCACCATCTGTCGACCGGAGGAGCGCCACCATCGCGCTGGCGGCGCGTCTGTCGCC 2092
QY 1967 TCTCCGACCGAGAGGAGGATTTCAACGAGACCGTGTGAAGGCCCGCGCATGTACCC 2026
      |||||
Db 2093 TGTCCGACCCCGACGAGAGGTCCGCGAATGCTCTCTCAAGGCGCAGACCACTTCGCC 2152
QY 2027 CCCT 2030
      |||||
Db 2153 CCCT 2156

RESULT 3
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Mutasynthesis
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ. ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Query Match 22.3%; Score 459; DB 3; Length 4496;
Best Local Similarity 60.5%; Pred. No. 4.2e-67;
Matches 899; Conservative 0; Mismatches 495; Indels 93; Gaps 5;

QY 2 TGGGACGCTTGTGATGACACATACGACTCTTTCACCCCACTGTTCAGTACATCG 61
Db 3028 TGGGAACCTGTGTATGACACATACGACTCTTTCACCTTTCAGATGCTGG 3087
QY 62 GCGAGGCAACGCGGCAACCCCGTCTGTCGTCGCAACGAGCGGACTGTGCGGGCTG 121
      |||||
Db 3088 CCGAGGGAACGCGCGCGCTCCGCTGCTCGTCCGCAACGAGCAACCCGACCTGGCAG 3147
QY 122 CCGTCAAG-----GACTTTCAGCGCATGCTGTGTCGCGGCGCCGCGAGCCGCGAC 175
      |||||
Db 3148 CCGTGGCGCGGAGCACTTTCAGCAACGTCGTCTCAACCCGCGCCGCGGCAACCCGCCA 3207
QY 176 GGGAGCGGACCTTTCGAAATCAAGCCGCGCGGCGATCAACCGACAGCGGCTTCGCTCG 235
      |||||
Db 3208 CCGACACCGACCTGCGGCTCAGCGCGCGGATGATCAACCGAATGGGACTCCCTGCTCG 3267
QY 236 GCGTCTGCTCTGCGGCAACGAGGATCGCCGAGCTCTTTCGCGGAAACGTCGCGCTCGCC 295
      |||||
Db 3268 GGGTGTCTGCGGCAACGAGGCTCTGTCTGCTCTGCGCGCGCGCGCTGTCTCAGCAGC 3327
QY 296 CCGAACCCATGCAACGCGCGGAGTCTCCGAGGTGCGGCAACCGCGAGAGAGTCTTCGCG 355
      |||||
Db 3328 CCGAACCTTTCAACGCGCGGACACGAGCAATCCGCCCAAGAGGCGCTGTTGCGCA 3387
QY 356 GCGTCTGCGCGCTTCAACCGCGGTGCGTACCACTCTGCGCGGCAACCTCCCG 415
      |||||
Db 3388 ACAATCCCTCCCGCTGACCGTGTGCTGCTACCACTGACCTGCGCAACTGCGCG 3447
```

```
QY 416 ACGAGCTCGAACCCCTCGCTGAGAGGAGCAACGAGGATCTCATAGGCTCTGCGGCAACCG 475
      |||||
Db 3448 CCGACCTGCGGCGCAACCGGCCCAACCGCGGAGCGGAGTGAATGGCCGTGCGCCACCGCC 3507
QY 476 AGAAGCGCTGTGGGCGCTTCAGTTTCACACCGGAGTTCATTCGCGACGCACTTCGCGCG 535
      |||||
Db 3508 ACTGTCCCGCTTCGCGCGGAGTTCACCCGAAATCGATTCAGACGCGCAACCGGCAAC 3567
QY 536 AGATCATGGCCAACTTCGCGGAGCT-----GTTACGAACTTCAGTGGCGCGCTTCGACGTGCTCGGAGCG 560
      |||||
Db 3568 GGAATGCTCGCCAACTTCGCGGAGCTGTCTCTGCGCGCGCGCGGCAACCGCCCGGCA 3627
QY 561 -----CGCCCTGCGCCACACCGGAGCAACGAGCGCAACGAGGCGCGACTCC 605
      |||||
Db 3628 CCGAAGCATACCCGCAACCGGACCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3687
QY 606 -----GTACGAACTTCAGTGGCGCGCTTCGACGTGCTCGGAGCG 646
      |||||
Db 3688 CGTCCGCGCGGATGGGAGATACCGGCTGCATGTGCGCAGGATGCTGCTGCGCGAGC 3747
QY 647 CCGAAGGATAGCGCGCGGCTGCTGCGCGGAGGAGCAACGTTCTGCTGAGCA 706
      |||||
Db 3748 CGAATCCGCGTTTCAACCGCTGTTGCGCGAGCGCCCGGCTCGGTTCTGAGTCAAGCA 3807
QY 707 GCTTCGCTCTGAAAGCGCGCTGCGGCTTCTCTCTGCGCGAGCAACGCGGCGCGCTG 766
      |||||
Db 3808 GCGGCTGCGAGCGGCGCTGCGCGCTTCACTTCTGCGCGCGCGCGCGCGCGCGCTG 3867
QY 767 CCGAGTACTCTACCTACCGCGCTGCGAGAGGCTGTCTCTGCTCGCGGCTCGACGAGCA 826
      |||||
Db 3868 GCGAACAATCACTACGACGCTGCGAGCGGCGGTGCGCTCAAGACGAGTTAAGGG 3927
QY 827 CCAGACCGCGGAGCGGCGCGCTTCTTCAATCTGAGGAGGAGCACTGCAACGCGAGC 886
      |||||
Db 3928 GCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3984
QY 887 GGGTCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 946
      |||||
Db 3985 GCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4044
QY 947 GCTACGAGCTGAAGGCGGAGCAACCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGAGC 1006
      |||||
Db 4045 GCTACGAGCAACGAGGCGGAGCAACCGGCGGCGGAGAGCGCGCGCGCGCGCGCGCGAGC 4104
QY 1007 CCGGTTCTCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1066
      |||||
Db 4105 GCGCTTCAATGTTGCGCGAGCGGAGTCTGCGCTGACCAAGAGAGGAGGCTGCTAC 4164
QY 1067 TGTGCGCTCTGACCGCGCGGCGCAACGAGCGGCGCGCGCGCTGCTGCGGAGAGCG 1126
      |||||
Db 4165 TCTGTGCACTGAGGAGGAGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 4224
QY 1127 CCGAGACCTTCAACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAGT 1186
      |||||
Db 4225 CCGAGCGCGCGCGGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4284
QY 1187 TCGGATTCGCGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1246
      |||||
Db 4285 ACACCAATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4329
QY 1247 ACGCTACTCAAGGCGCATGAGTGTCTCAAGAGATTCGCAACGCGGAGTGTAG 1306
      |||||
Db 4330 CCGGCTACGCGGAGATGATGAGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 4389
QY 1307 AGATGCTCTGACCAACATGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACT 1366
      |||||
Db 4390 AGGTGTCTGAGCAACATGTCTCGGAGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCTTAC 4449
QY 1367 CCGCGTGGCGCGCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1413
      |||||
Db 4450 GCGCGCTGCGACCGGTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4496
```

RESULT 4
US-09-987-614A-6
; Sequence 6, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMES-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR FILING DATE: 1997-03-20
; SOFTWARE: Patentin Ver. 2.0
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-6

Query Match 22.3% Score 459; DB 4; Length 4496;
Best Local Similarity 60.5%; Pred. No. 4.2e-67;
Matches 899; Conservative 0; Mismatches 495; Indels 93; Gaps 5;

QY 2 TGGCAGCGCTTGTGATGACAACTAGACCTGTTCAACCACTGTTCCAGTACATCG 61
DB 3028 TGGGAACCTGCGATGACAACTAGACCTGTTCAACCACTGTTCCAGTACATCG 3087
QY 62 GCGAGCCACCGGGCAACCCCGTGTGCTGCCAAGCAGCGGAGCTGCGGCTCG 121
DB 3088 CCGAGGTGAACCGCGCGCTCCCTGCTGCTCCGAAACGACACCCGCACTGGCAGG 3147
QY 122 CGGTGAG-----GACTTGAAGCGATGCTGCTGCTCCCGGGCCCGGAGAGCCGAC 175
DB 3148 CCTGAGCGCGGCGGAGCTTGAACAGTGTGCTGCTCACCAGGCTCCGCGACCCGCGCA 3207
QY 176 GGGAAAGGACTTGGAAATCAGCGCGGCGGATCAACGACGAGGCTGCGGCTCG 235
DB 3208 CCGAACCGACCTGGGCGCTGAGCGCGGAGTATCAACGAAATGGAGCTGCGGCTCG 3267
QY 236 GGGTGTGCTGCGGCAACGAGGAGTGCCTTTCGCGGAAACCTGCGGCTGCGG 295
DB 3268 GGGTGTGCTGCGGCAACGAGGCTGCTGCTGCTGCTGCGGCGCGCGCTGTCAGCAG 3327
QY 296 CCGAACCTTGAACCGCGCGGAGTCTCCGAGGTGGGACACCGCGGAGAGCTTTCCGG 355
DB 3328 CCGAACCTTTCACGCGCGGACAGGACATTCGCGACGAGGAGGAGCTGTTTGGCA 3387
QY 356 GCTTCCCTGCGGCTTACCGCGGCTGAGTACCACTCCGCGGACCGGACCTGCGG 415
DB 3388 ACATCCCTTCCCGCTGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3447
QY 416 ACGAGCTGAACCCCTGCTGCTGAGGAGGAGCGAGGAGTGTGATGAGGCTGCGGACCG 475
DB 3448 CCGACTGTGCGGCGCACCGGCAACGCGGAGGAGTGTGAGGCTGCGGACCGGCG 3507
QY 476 AGAAGCGCTGTGGGCGCTTCAATTCACCGGAGTTCATCGGACGACTTGGCGGAG 535
DB 3508 ACCTGTCCCGCTTGGGCTGTGAGATTCACCCGAAATGATCAACGAGCAACGCGGAC 3567
QY 536 AGATCATGGCAACTTCCGAGACT----- 560

DB 3568 GATGCTCGCAACTTCGCGGACCTGTCCCTGTGCGCGGCGGCGACCGGCCCCCGCACA 3627
QY 561 -----CGCTTGGCCCAACCGGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 605
DB 3628 CCGAACGATACCGGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3687
QY 606 -----GTACGAACCTCAACGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 646
DB 3688 CGTCCGCGCGGCGGAGTACCGGCTGATGTGTGCGAGGTGCGCTGCTGCTGCGGAG 3747
QY 647 CCGAAGAGTACCGCGGCGGCTGCTGCGGCGGAGGCAACAGTTCGAGTGAACGCA 706
DB 3748 CCGAGCGCGGCTTACCGCGGCTGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 3807
QY 707 GCTTCCCTTCCGAAGCGGCTGCGGCTTCTCTTCTGCGGCGGCGGCGGCGGCGGCGG 766
DB 3808 GCGGCTTCCGAGCGGCGGCTGCGGCGGCTTCACTTCTGCGGCGGCGGCGGCGGCGG 3867
QY 767 CCGAGTACCTTCACTGCGGCGGCTGCGGAGGCGGCTGCTGCGGCGGCGGCGGCGGCGG 826
DB 3868 GCGAACAGTACACTGACGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3927
QY 827 CCAAGACCGGAGCGCGGCGGCTTCTTCACTGAGGAGGAGGAGGAGGAGGAGGAGG 886
DB 3928 GCGAGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3984
QY 887 GGGTCCCGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 946
DB 3985 GCGGCTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4044
QY 947 GCTAGAGCTGAAGGCGGAGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1006
DB 4045 GCTAGAGCACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4104
QY 1007 CCGGCTTCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1066
DB 4105 GGGCTTATGTTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4164
QY 1067 TGTGCGCTTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1126
DB 4165 TCTTGGCACTGAGCAGCACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4224
QY 1127 CCGAAGCCTTCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1186
DB 4225 CCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4284
QY 1187 TCGGATCCCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1246
DB 4285 ACGACCACTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4329
QY 1247 ACGCTTACTTAAGGCGATGAGTGTCTTAAGAGATTCGCAACGCGGAGTGTGATG 1306
DB 4330 CCGCTTACCGGGAATGTTGAGGATGCGCGGCTGATTCACGAGCGGCGGAGCTAG 4389
QY 1307 AGATGCTTGAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1366
DB 4390 AGTGTGCTTGAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4449
QY 1367 CCGGCTGAGCGCATACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1413
DB 4450 GGGCGCTGCGCACCGTTCAGCCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4496

RESULT 5
US-08-765-907A-1
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMES-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel

```
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHÉ, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Multisynthesis
; CURRENT APPLICATION NUMBER: US/08/765,907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

Query Match      17.7%; Score 365.4; DB 3; Length 2888;
Best Local Similarity 74.2%; Pred. No. 1e-51;
Matches 462; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1408 CTCGAGTCCCGCACTGTCGTCGAGCGCTCCGCGGAGGTTCTCAGATCGGC 1467
DB 1 CTCGAGTCCCGCGGCGCACCGTCTCAGCTCTCTCAGCGATTCTGCGCATCGGC 60
QY 1468 GCCGACGCGCGCGCTCGAGTCCAGCCATCAAGCCGAGCCGCCCGCGCGGCGCACCGCG 1527
DB 61 GCGGAGCGGCTGGCGGAGTCCAAACCATTAAGGACCGCGCCCGCGCGCGCGCGCGCGCG 120
QY 1528 GAGGAGAGCAGAGCGGCTCCGCGCGCACTGCGCGCGCGGAGAGAGACCGGCGCGAGAAC 1587
DB 121 GCCCAGAGACCGCGCGTCAAGGCTCTCCGCGCGCGCGAGAGACCGCAGCGAGAAC 180
QY 1588 CTGATGATGTCGACCTGTCGCGCAAGACCTCAAGAGTCTGCGCGATCGGCTCTGTC 1647
DB 181 CTATGATGTCGACCTGTCGCGCAAGACCTGCGCGAGTCTGCGCATCGCTCTGTC 240
QY 1648 CACGTGCCCCGCGCTCTTCGAGTGAAGACCTTACGCGCGCGTGAAGACCTGATGTCGAC 1707
DB 241 CACGTACCGGCGCTGTTTCGAGTGAAGACCTTACGCGCGCATCGACGCTCTGTCGACG 300
QY 1708 ATCCGAGAGCGCTGCGCGCGCGCGACCAAGACCGCGCGCTGCGTACGCGCGCTTCC 1767
DB 301 GTCGCGCGCGCGCTGCGCGCGCGCGTCTCCGCGCGCGCGGTACGCGCGCGCTTCC 360
QY 1768 GCGGCGCGCATGACCGCGCGCGCGCGCAAGAGCGCGCATGAGATCCAGCGCGCTGAG 1827
DB 361 GCGGCGCGCATGACCGCGCGCGCGCGCGCAAGAGTCCGACATGATTCAGACCGCTGAG 420
QY 1828 GAAGCGCGCGCGCGCTCTTACTCCGAGCGCGCTGAGATGTTCCGCTTCAAGCGCGCG 1887
DB 421 AAGGCGCGCGCGCGCTGTTACTCGGCGCGCTGAGGCTGAGTCTTCCGCTTCAAGCGCG 480
QY 1888 GACCTCAGCATGTCATTCGCGACCATGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 1947
DB 481 GACCTCAGCATGTCATTCGCGACCATGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 1948 GCGGCGCGCGATGTCGTCCTCTCCGACGAGAGAGAGTTCAGCGAGCGCGTGAAG 2007
DB 541 GCGGCGCGCGATGTCGTCCTCTCTCCGACGAGAGAGAGTTCAGCGAGCGCGTGAAG 600
QY 2008 GCCCGCGCGATGTCACCGCGCT 2030
DB 601 GCGCAGACCACTCTCGCGCGCT 623

RESULT 6
US-09-987-614A-1
; Sequence 1, Application US/0987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
```

```
; APPLICANT: THIBAULT, Denis
; APPLICANT: BAMES-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHÉ, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Multisynthesis
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Query Match      17.7%; Score 365.4; DB 4; Length 2888;
Best Local Similarity 74.2%; Pred. No. 1e-51;
Matches 462; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 1408 CTCGAGTCCCGCACTGTCGTCGAGCGCTCCGCGGAGGTTCTCAGATCGGC 1467
DB 1 CTCGAGTCCCGCGGCGCACCGTCTCAGCTCTCTCAGCGATTCTGCGCATCGGC 60
QY 1468 GCCGACGCGCGCGCTCGAGTCCAGCCATCAAGGAGACCGCGCGCGCGCGCGCGCGCG 1527
DB 61 GCGGAGCGGCTGGCGGAGTCCAAACCATTAAGGACCGCGCCCGCGCGCGCGCGCGCGCG 120
QY 1528 GAGGAGAGCAGAGCGGCTCCGCGCGCACTGCGCGCGCGGAGAGAGACCGGCGCGAGAAC 1587
DB 121 GCCCAGAGACCGCGCGTCAAGGCTCTCCGCGCGCGCGAGAGAGACCGCAGCGAGAAC 180
QY 1588 CTGATGATGTCGACCTGTCGCGCAAGACCTTCAAGAGTCTGCGCGATCGGCTCTGTC 1647
DB 181 CTATGATGTCGACCTGTCGCGCAAGACCTTCAAGAGTCTGCGCATCGCTCTGTC 240
QY 1648 CACGTGCCCCGCGCTCTTCGAGTGAAGACCTTACGCGCGCGTGAAGACCTGATGTCGAC 1707
DB 241 CACGTACCGGCGCTGTTTCGAGTGAAGACCTTACGCGCGCATCGACGCTCTGTCGACG 300
QY 1708 ATCCGAGAGCGCTGCGCGCGCGCGACCAAGACCGCGCGCTGCGTACGCGCGCTTCC 1767
DB 301 GTCGCGCGCGCGCTGCGCGCGCGCGTCTCCGCGCGCGCGGTACGCGCGCGCTTCC 360
QY 1768 GCGGCGCGCATGACCGCGCGCGCGCGCAAGAGCGCGCATGAGATCCAGCGCGCTGAG 1827
DB 361 GCGGCGCGCATGACCGCGCGCGCGCGCGCAAGAGTCCGACATGATTCAGACCGCTGAG 420
QY 1828 GAAGCGCGCGCGCGCTCTTACTCCGAGCGCGCTGAGATGTTCCGCTTCAAGCGCGCG 1887
DB 421 AAGGCGCGCGCGCGCTGTTACTCGGCGCGCTGAGGCTGAGTCTTCCGCTTCAAGCGCG 480
QY 1888 GACCTCAGCATGTCATTCGCGACCATGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 1947
DB 481 GACCTCAGCATGTCATTCGCGACCATGTCGTCGCGCGCGCGCGCGCGCGCGCGCG 540
QY 1948 GCGGCGCGCGATGTCGTCCTCTCTCCGACGAGAGAGAGTTCAGCGAGCGCGTGAAG 2007
DB 541 GCGGCGCGCGATGTCGTCCTCTCTCTCCGACGAGAGAGAGTTCAGCGAGCGCGTGAAG 600
QY 2008 GCCCGCGCGATGTCACCGCGCT 2030
DB 601 GCGCAGACCACTCTCGCGCGCT 623
```

RESULT 7
US-08-403-852D-9
Sequence 9, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyer, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 61..645
OTHER INFORMATION: /product= "gene papa"
US-08-403-852D-9
Query Match 13.2%; Score 272.2; DB 2; Length 645;
Best Local Similarity 68.7%; Pred. No. 2.2e-36;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;
QY 2 TGGGACGCTTTCGATGAGAACTAGACACTGTTACCCCAACCTGTTCACTACATCG 61
DB 71 TGGAAACCTCGATGATGAGAACTAGACACTGTTACCTTCAACCTCTTCACATGCTGG 130
QY 62 GCGAGGCCAACCGGCAACCCCGCTGCTGCGCAACGAGCGGCACTGGTCCGGCTGC 121

DB 131 CCGAGGTGAAGCGGCGCGCTCCGCTCGTCCGCAACGACGACACCCGCACTGGCAGG 190
QY 122 CCGTCGAG-----GACTTGAGAGCCATTCCTGCTGTCCTCCCGGCGCCCGGCAAGCCCGAC 175
DB 191 CCGTGGCGCGGCGGAGCTTGAGCAACGTCGCTGCTCACCAGCGCCCGGCGACCCCGCA 250
QY 176 GGGAAACGGACTTCGGAATCAGCCCGCGGCGATACGACGACGAGCGCTGCTGCTCG 235
DB 251 CCGAACACGACCTGAGGCGCTAGCCCGGAGTATCAGCAATGAGACCTCCGCTGCTCG 310
QY 236 GCGTCTGCTCGGCAACGAGGAGATGCGCCAGCTCTTTCGCGGAAACGTCGAGCTCGCC 295
DB 311 GGGTGTCTCTGGGCAACGAGCCCTGCTGCTGCTCGCGCGCGCGGCTGCTCAGACAC 370
QY 296 CCGAACCCATGACAGCGCGGCTCTCGAGGTGGGACACCGGCGAGAGACTTTCGCGG 355
DB 371 CCGAACCTTTTCACGCGCGGCGACGAGGACATTCGCGCACGAGCGGCGGCTGTCGCA 430
QY 356 GCGTCCCTCGCGGCTTACCGCGCGTGGCTACCACTCCCTGCGCGCCACCGACCTCCCG 415
DB 431 ACATTCCTCTCCGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490
QY 416 ACGAGCTGAAACCTCTGCTGAGGAGGACGAGCGGCTGCTGATGAGGCTGCGGACCGCG 475
DB 491 CCGACCTGGCGGCGCACCGCCACACCGCGGACGAGTGGCTGCGCCACCGCG 550
QY 476 AGAAGCGCTGTGGGCGCTCACTTCCACCGGAGTCCATGCGAGGACTTTCGCGCGG 535
DB 551 ACCTGCCCGCTTTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 610
QY 536 AGATCATTGCGCACTTCGCGGACCTGCGCT 566
DB 611 GGAATGCTGCGCACTTCGCGGACCTGCTGCT 641
RESULT 8
US-08-510-646B-9
Sequence 9, Application US/08510646B
Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 61..645
OTHER INFORMATION: /product= "gene papa"
US-08-510-646B-9

Query Match 13.2%; Score 272.2; DB 3; Length 645;
Best Local Similarity 68.7%; Pred. No. 2.2e-36;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;
Db 2 TGGGACGGCTTGATGACAACTAGCACTGCTTCAACCCCAACCTGTTCCAGTACATCG 61
71 TGGGAACCTGCTGATGACAACTAGCACTGCTTCAACCACTCTTCCAGATGCTGG 130
Qy 62 GCGAGGCAACCGGGCAACCCCGCTGCTGTCGCAACGACGCGACTGTCGGGCTGC 121
131 CCGAGGTGAAACGGCGCGCTCGCTGCTGCTGCGCAACGACGACCCGCACTGGACAG 190
Db 122 CCGTCTGAG-----GACTTGCAGCCGATGCTGTCCTCCCGGGCCCGGACGCCGAC 175
191 CCTGGGCGCGGCGGCTGATGACAAAGTCTGCTTCAACCGGCGCGGCGCACCCGCCCA 250
Qy 176 GGAACGGGACTTCCGAATAGCCCGCGGCGGATCAACGACGAGGCTGCGCTCTCG 235
251 CCGAACCGGACTGCGGCTGAGCCGCGGGTGAATCAACGAAATGGGACCTGCGCTCTCG 310
Db 236 GCGTCTGCTCGGCGGCAACGAGGATCGCCAGCTCTTCCGCGGAACTGCGGCTGCGCC 295
311 GGGTGTGCTGCGGCGGCAACGAGGCTGCTGCTGCTGCGGCGCGCGCTGCTGCTCAAG 370
Qy 296 CCGAACCGGATGACGCGCGGGGTCTCGAGGTGCGGCAACCGGCGAGAGAGCTTTCGG 355
371 CCGAACCTTTCACGCGCGGCAACGAGGATCGGCGGCGAGGAGGCGCTGTTGCGGA 430
Db 356 GCGTCTGCTCGGCGGCTTCAACGCGGCTGAGTACCACTTCCGCGGCGGCACTTCCCG 415
431 ACATCCCTTCCGCGGCTGAGCGGTGCTGATCACTGCTGAGCCGTCGCGCAACTGCGCG 490
Qy 416 ACAGCTGCAACCCCTGCTGAGAGGACGAGCGGGGTCTCATGGGCTTCCGCGGCGCGG 475
491 CCGAAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550
Db 476 AGAAGCGGCTGTGGGGGCTCAAGTTCACCGGAGTCCATCGGACGAGCACTTTCGGCG 535
551 ACCTGCGCGGCTTCCGGGCTGAGTTCACCGGAGTCCATCGGAGCAAGGAGGCGGCGG 610
Qy 536 AGATCATGGCACTTTCGGGAGCTTCCGCTT 566
611 GGATGCTCGGCAACTTCCGCGGAGCTTGCCT 641

RESULT 9
US-09-231-818-9
Sequence 9, Application US/09231818
Patent No. 6171846
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Blanc, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
TITLE OF INVENTION: Polypeptides Involved in The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pinegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/231,818
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 645 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 61..645
OTHER INFORMATION: /product= "gene papa"
US-09-231-818-9
Query Match 13.2%; Score 272.2; DB 3; Length 645;
Best Local Similarity 68.7%; Pred. No. 2.2e-36;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;
Db 2 TGGGACGGCTTGATGACAACTAGCACTGCTTCAACCCCAACCTGTTCCAGTACATCG 61
71 TGGGAACCTGCTGATGACAACTAGCACTGCTTCAACCACTCTTCCAGATGCTGG 130

Accession	Sequence	Position
Db	1074 CGTCAACGGCGAACTTGGCGCCCGGGGCAGAGACGCCCTTCCACTGCTGGAAAGGACGTTTCC	1133
QY	1757 CGGCGGCTTCATGATGACCGGCGCGGCCCAAGAGCGACCATGGAGATCATGACCCGCTTGA	1826
Db	1134 CGGCGGCTTCGATCACCGGCGCGCCGAGATTGGCGCATGGCAGATCATGACGAACCTGA	1193
QY	1827 GGAAGGCCCCCGGGGCGCTTACTCCGGGCGCTCGAATGATTTCCGCTTCAAGCGGCGCCG	1886
Db	1194 ACCGAGCCGACGGGCGATCTACTGGGAGACCTGTTTCACTTCGACGTCGCGCGCGAGAT	1253
QY	1887 CGACCTCAGACATCTCATTCGCGACCATCGTGTGCTGCGACGCGCGACGCGAGATTCCGGGT	1946
Db	1254 GGAACAGCTGCATTCGCGCATTCGCGACCTGCTGTCAGGAACGGCCAGTCAAGTTGCTGGGG	1313
QY	1947 CGGCGGGGCGATTCGTGTCCCTTCGACCAAGAGAGGAGATTTCACCGAGACCTGTGTAA	2006
Db	1314 CGGCGCGCGCATTCGTGCGCGCATCGCATCTGGGAGAGACGATGACAGAAACCTTCGACAA	1373
QY	2007 GGGCGGCGCATTCGTCACCGGCGCTTCGACGAGCGCGCCGCTGGCGGGCGGCC	2056
Db	1314 GATCCGGGTGCTCTGGAACCTCTGGAAGGAGTGGCCGGGACAGGTTCCC	1423

```

RESULT 13
US-09-252-991A-13943
; Sequence 13943, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Matt J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13943
; LENGTH: 816
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13943

```

Query Match	12.6%	Score 260	DB 4	Length 816
Best Local Similarity	61.4%	Pred. No. 2.2e-34		
Matches 435	Conservative 0	Mismatches 270	Indels 3	Gaps 1
QY	1349	CGGCCCTGCGCCTCTACTCTCGCGCTGCGCGCCATCAGCCCGCTCCGTACCGGCCCTGC	1408	
Db	30	CGCGCTGGCCGAGCCTATCGCGCCCTCGCGCGAGGCGTGGCGGACCGCCGTCTCCGGCTACC	89	
QY	1409	TCGAGTTCCTCCGAACTGTCGCTGCTGAGCGCCTTGCCCGAGCGTTCCTCAGATCGAGC	1468	
Db	90	TGCGCATGGCCCGGACGGCGCGATCTCTCAGCCTGTGCGCGGAGCGCTTCTCAAGCTCGACA	149	
QY	1469	CCGACGGCGCGCTCGAGTCCAAAGCCCATCAAGGGGACCCCGCCCGCGCGCACCCCGG	1528	
Db	150	---AGGGCAGGTGGAAACCCCGCGCATCAAGGGACACC CGCCCGCGGACAGACCCCG	206	
QY	1529	AGGAGGACGAGCGGCTCCGCGCCGACCTGCGCGCGCCGGGAGAGAGACCGGGCCGAGAAC	1588	
Db	207	AGGAGGACATGAGCGCTGGCGCGCGTGCCTGTGGCGACGCCCAAGACCTCGCGGAAAAAC	266	
QY	1589	TGATGATGTGACCTGTGCTCCGACACGACTCTCAACAGCGTCTGGCGCATGGCTCCGCTCC	1648	
Db	267	TGATGATGTGACCTGTGCTGCGACACGACATCGACGAGTTGCCAACTCGGCGAGCTTAC	326	
QY	1649	ACGTGCCCCGGCTCTTGAGAGGTGAGAGCATAGCGCCCGCTGACACAGCTGTGTTCAGCA	1708	
Db	327	GGGTACCGGAGACTGTTGCCCTCGGAAAGCTATCCCAAGTGCATCTACCTGGTGAAGCAGC	386	

Qy	1709	TCGGGGACGGCTGCGGCGCGGGACACGACACCGCGCGTCGCTACGGCGCGCCCTTCCCG	1768
Db	387	TCACCGGGAACTGAGCCCGCGGCAAGGACGGCTCTGACTCGCTGGAAAGGACAGCTTCCCG	446
Qy	1769	GGCGCTCCATACCGCGCGCGGCCCAAGAGCGCACATGAGAGATCATGACCGCGCTGGAG	1828
Db	447	GGCGCTGATACCGCGCGCGGCCCAAGATTGCGCCCATGCAATATATGACGAACTGGAAC	506
Qy	1829	AAGCGCCCGGGGCGTCTACTCCGGGGCGCTCGGATGGTTGCGCCCTCAGCGCGCGCGCG	1888
Db	507	CGAGCCGACGGCGGATCTACTGGGGGAGCGCTGTTCTACTCTGACGTGCGGGGCGAGATGG	566
Qy	1889	ACCTCAGCATGTATCCGACCATCTGCTGCGCCGACGCGCCAGGCGGAGTTGCGCGTCG	1948
Db	567	ACAGCTGATCGCATCCGACCTCTGCTGCTCAAGAACGGCCAGGTCAGTTGCTGGGGCG	626
Qy	1949	GCGGGGCGATGCTGCTCTCCGACCAAGAGAGAGATTCACCGAAACCGTGGTAAGG	2008
Db	627	GCGCGGCGATGTGCGCGACTCGACTCGAGGAGGACGAGTACAGGAAACCTTGGACAAAG	686
Qy	2009	CCCGCGCAATGTACCGCCCTCGACGGACGCGCGCGGGGCGCGCC	2056
Db	687	TTCGGGTGCTCTGAAACCTCGAAGGAGATGCGCCGGGACAGCGTCCC	734

```

RESULT 14
US-09-902-540-5760
: Sequence 5760, Application US/09902540
: Patent No. 6833417
: GENERAL INFORMATION:
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: APPLICANT: Wiegand, Roger C.
: TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
: FILE REFERENCE: 38-10(1584)B
: CURRENT APPLICATION NUMBER: US/09/902,540
: CURRENT FILING DATE: 2001-07-10
: PRIOR APPLICATION NUMBER: 60/217,883
: PRIOR FILING DATE: 2000-07-10
: NUMBER OF SEQ ID NOS: 16825
: SEQ ID NO 5760
: LENGTH: 1473
: TYPE: DNA
: ORGANISM: Myxococcus xanthus
US-09-902-540-5760

```

Query Match	11.3%	Score 232.6	DB 4	Length 1473
Best Local Similarity	49.4%	Pred. No. 7.1e-30		
Matches	663	Conservative	0	Mismatches 674; Indels 6; Gaps 2
Qy	719	AAGCGCCTCGCGCTTCTCTTCCCTCGGCAAGACCGCGGCGCCCTCGCCAGTACTCTCA	778	
Db	137	ACGGGGGGAGAGGGTTTGGCGCGCTACAGACCAATCGGCGCGAGACCCGGGGCCGGCTGC	196	
Qy	779	CTTACCGCGTGCAGACGGCGGTGTCTTCCTCGCGGCTCGACGAGCAACAGACCCGGA	838	
Db	197	GCTTGGACACGGAAAGGTGCGCACTTTGGGCGCGGCTCGCGGAGAGAGCGGGACGGCA	256	
Qy	839	CGGGGCGCCCTTTCTTCACTACTGGAGAGACGCTCGAAACGCCGACGGGTCCCTCGTG	898	
Db	257	ATCCCTTGAGATGTCCTCCGCGACGCTCTGGCGGAGCTGGCGGTGACCCGCTCCCTGGAG	316	
Qy	899	CCCCGAACTGCGCCTTGAGATTCAACTCGGCTACGTACGCTACTCTGGGCTACGAGCTGA	958	
Db	317	AGGCCCCCTTCTTGCGGGGCTCTGTGGGTACATAGGCTTACAACTGTCTTCTGTGGCTGG	376	
Qy	959	AGGCGAGACCAACCGGCGACCCCGCGACCGGTCCCGCAACCCGACGCGCGTTTCTCT	1018	
Db	377	AGCCCTCGTGGCGGACCGGACCCCAAGCACAATCTTCTTCCGGACTTCGGAATGGCTGG	436	
Qy	1019	TGCGCGACCGGCGCCTCGCCCTCGACCAACGAGAGGCTGCTACTCTTGCGGCGCTCG	1078	

Db	437	TTGTGCGAAGGACTTTCCTCACGCAACGACACCCCGACCCGGAGCGCTCAAGGCCACCGCATTTG	1138
QY	1079	AACCGCGGGGCGCACGACGACGCGCGCCCGCGCTGGCTGGCGGAGACGCGCCGAGACCTTCA	1138
Db	497	CCCGTCGCTCGCTGACGCGGACGCGCGTGGCGCGCGCTGAAGGACGCGGAGGGGCGCGGC	556
QY	1139	CCGCGCTTGGCGCTTCGCGCGCCCGCGCGGACCGACCCCTCGCATGTGCTTTCGGGATCCCGG	1198
Db	557	AGGCCATATGGCGGATG---GTTGTTGAAACCACTGCGCACCAAGAGCGCTTATGCCCGAG	612
QY	1189	AGGCGCGCGCGGCTTTCGCGCCCTTCGCGCGCGCGCGCGCGACGACGACAAAGGACGCTACCTCA	1258
Db	613	CCGCGCATGTC--GTGGGAGACCGCGCGCTTACCGCGTGTGGAGACCGCGCGGATGAGT	670
QY	1259	AGCGCATGACGAGTGCCTCAAGAGATCCGCAACGCGCACTGCTATCGATCTTGCTGA	1318
Db	671	CCGCGGTGGCTCAGCGCGGACGAGGATGATCCGCGCGGAGACATCTTCCAGGTCCGTCG	730
QY	1319	CCACATATGATCAACGCGCGGACCGAGGCGGACGCGCTTCGCGCTTACTTCGCGCTGGCG	1378
Db	731	CGCGCGGCTTGCATCCCGGGGCGGCGCGCGCGCTGTCCCTTACCGCGCGCTGGCGGC	790
QY	1379	CCATGAGCGCGTCCCGTATCGGCGCGCTCGCTGAGTTCCCGGAATGTCGGTGTGAGCG	1438
Db	791	GAGTAAATCAATCGCGCTGATCTTTTCTCTGTGAGATCTGGGAGAGCCGCGCGCTGTGG	850
QY	1439	CCTCGCGCGAGGAGTTCCTTCAACGATCGGCGCGCGACGCGCGCGTGCAGTCCAAAGCCATCA	1498
Db	851	GCGCGTCCCGCGAAGTTGCTGGATGCAAGTGTCCGACGAGGAGATGTGATGTGGCGCCATCG	910
QY	1499	AGGGAGCCCGCCCGCGGGCGGACCGCGGAGAGAGCAGACGGCTCCGCGCGCACTTGG	1558
Db	911	CCGCGACCCCGCGCGCTGTGCTGTCCGAGCGCGAGGACCTTGAGCTTGAGAAAGAGCTGC	970
QY	1559	CCGCGCGGGAGAAAGAACCGCGCGCGAGAACCTGATGATTTGTGACTGTGTCCGAAACGACC	1618
Db	971	TGCGCGAGAGAAAGAGCTTCGCGGACCATATGCTGTGACCTTGGGGGCGCAACGACG	1030
QY	1619	TCAACAGCGTCTGACCGATTCGGGCTTCGTCACAGTGCACCCCGGCTCTTGCAGGTGAGAGACT	1678
Db	1031	TGGCGCGCGTGGCGCGCGCTGGCTGTGGTGTGCGGTGAGGAATGATGCTCATCGAGCGCT	1090
QY	1679	ACGCGCGCGTGCACCAAGCTGTGTGACCATTCGGGAGACGGCTGCGCGCCCGGACCAACA	1738
Db	1091	ACAGCGCATGTATGACATCGTGTGTGCGAGTGTGCGGCGCAAGCTCGACCGCAAGTACGACG	1150
QY	1739	CCGCGCGCTGCGTACGCGCGCGCTTTCGCCGCGGCTCATATGACCGCGCGCCCAAGAAGC	1798
Db	1151	CGCTGGATGCGCTGGCGACCAACCTTTCGCCGCGGACGATGTGCGGCGCGGCCCAAGATTTC	1210
QY	1799	GCACCATGAGATTCATTCGACGCGCTCGAGAGAAAGGCCCCCGGAGCGTACTTCGCGGCGCGC	1858
Db	1211	GCGCGATCAATATCATCGACAGATTTGAGAGTTCACGCGCGCGGCGCTTATTCGCGCGCGG	1270
QY	1859	TGCGATGCTGCGCTTCACGCGCGCGCGCGCACTTCAGCATGTCATTCGCAACCATCTGTGC	1918
Db	1271	TGGGCTTACTGTCTTCTGCGGCAACGTGAGCATTTGGCTGTGGACCTTTCTTTCG	1330
QY	1919	TGGCGGACGCGCGCGGAGTTCGCGCTGCGCGGGGCGATGTGTCTCTTTCGACCAAG	1978
Db	1331	TGAGACGAGACCGGACGATGTGTGACCTTGGGCGCGGAGTGTGCGGCACTTCGCTGCGCT	1390
QY	1979	AGGAGGATTCACCGAGACCGGTGTAAAGGCGCGGCATGTATCAACGCGCTTCGACGCA	2038
Db	1391	CGAAGGAAGCGGACCAAGACGGAAGCAAGGCGGCGCATGACCGCGCGCTGCGCGCTGG	1450
QY	2039	GCGCGTGGCGGCGCGCGCATGA 2061	
Db	1451	CGCGGAGAGAGTGGCCGCTGA 1473	

```

US-09-902-540-1272/c
; Sequence 1272, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10115849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1272
;
; LENGTH: 54484
;
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1272

```

Query Match	11.3%	Score 232.6	DB 4	Length 54484
Best Local Similarity	49.4%	Prid. No. 7,5e-304	Mismatches 634	Indels 6
Matches	663	Conservative	0	Gaps 2
QY	719	AAGCGCGCTCGGCGCTTCTCTTCTCGGCGAGCAACCGCGGCCGCTCGCGAGTACTCA	778	
DB	49955	ACGGCGCGAGCGGCGGTTCCGCGCTTACAGCAATCGCGCGGAGCCCGCGCGCGTGC	49896	
QY	779	CTTACCGCGTGCAGCGAGCGTGTCTCCGTCGCGGACTCGACGAGCAACAGACCCGGA	838	
DB	49895	GCTTGACCGGAACGGGTGCACCTTGTGCGCGGCTCGCGAGGAGCGCGGAGCGCA	49836	
QY	839	CGGCGCGCGCGCTTCTTCACTACTCTGAGAGACAGTGCAGCGGACGCGGCTCCCGCTCG	898	
DB	49835	ATCCCTCGAGTGTGTCTCCGCACTCTGACGAGGTGCAGTGCACCGGCTCCGCTCTTGAG	49776	
QY	899	CCCCCGAATCGCCCTTTCGAGTTCAACTCGGCTACGTGAGTACCTCGGCTTACGAGCTGA	958	
DB	49775	AGGCGCGCTTCTCTCGCGGCGCTGTGGGTACATGAGGCTTCAACTGTCTTCTGTGGTGG	49716	
QY	959	AGGCGAGAACCAACCGGAGACCCCGCGCACCGGTCCCGGACCCCGACGCGCGTCTCT	1018	
DB	49715	AGCGCTCGTGTGCGGACCGGCAACCCAGGACATCTCTTCCCGGACTCGAATGGCTGG	49656	
QY	1019	TTCGCCAACCGCGCGCATTCGCTTGTGACACACAGGAAGGCTGTCTACTGTCTGGCGCTCG	1078	
DB	49655	TGTGCGAGGACTTCTGTGTACGACAGAACCCGGAACCGGAGCGCTCAAGGCGCAACCGGCATTG	49596	
QY	1079	ACCGCGGAGGCGCAACGAGCGGCGCGCGCTGTGAGGGAGACGCGCGGACCTTCA	1138	
DB	49595	CCGTCCTGTGCTGTGCACAGCGAGGTGTGGCGCGCTGTGAAGAACGCGAGGGCGCGGC	49536	
QY	1139	CGGCGCTGTGCGCGCGCGCGCGCGCGCGGAGCGAACCCCGCGCAATGTCTTTGGGATCCCG	1198	
DB	49535	AGGCGATGTGGGATATG---GTTGTGAACAACATCTGCACACGAGGCTATCGCGCGAG	49480	
QY	1199	AGGCGCGCGCGCGCTTGTGCGCGCTTGTGCGCGCGCGCGCGCAACGACGACTTACTTCA	1258	
DB	49479	CCGCGCATATGCG--GTGGGAGACGCGCGCGCTTACCGCGTGTGGGACCGCGCGGGGTATGAGT	49422	
QY	1259	AGCGCATTCGACGAGTGCCTTACAGAGAAATCCGCAACGGCGGAGTGTGACGAGTCTGCCGA	1318	
DB	49421	CGGCGGTGGCTCAGGCGCGAGAGTAAATCCGCGCGGAAACATTTTCCAGGTGTGCTTG	49362	
QY	1319	CCAACATGTTCACCGCGCGCACGAGGACACGCGCTGTGCGCTTACTTCGCGCTGCGCG	1378	
DB	49361	CGCGCGCTTTCGAAATCCCGGGGCGCGCGCGCGCGCTGTCTTACCGCGCGCTGCGGC	49302	
QY	1379	CCATCAGCCCCGTCCTGTACGCGCGCTGTCTGAGTTTCCCGAATGTGTGCTGACGC	1438	
DB	49301	GGGGAATCATCATGCGGTAATCTTTTCTGTGAGCTGTGGACGAGCGCGCGGCTGTGG	49242	

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 ; Search time 2023.38 Seconds
(without alignments)
7033.027 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061
Sequence: 1 atcgacgacgtcttcgatcgca.....ccgtgagcgaggcccgatga 2061

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 0%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
- 13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/prodata/1/pubpna/US10D_PUBCOMB.seq:*
- 17: /cgn2_6/prodata/1/pubpna/US10E_PUBCOMB.seq:*
- 18: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*
- 19: /cgn2_6/prodata/1/pubpna/US10F_PUBCOMB.seq:*
- 20: /cgn2_6/prodata/1/pubpna/US10H_PUBCOMB.seq:*
- 21: /cgn2_6/prodata/1/pubpna/US10I_PUBCOMB.seq:*
- 22: /cgn2_6/prodata/1/pubpna/US10I_NEW_PUB.seq:*
- 23: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq:*
- 24: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
- 25: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
- 26: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2061	100.0	2061	20	US-10-472-587-1
2	1773.8	86.1	3305	14	US-10-126-927-68
3	1772.2	86.0	12391	14	US-10-126-927-67
4	769.8	37.4	138203	21	US-10-819-386A-1
5	606.2	29.4	2196	15	US-10-156-761-1167
6	606.2	29.4	9025608	15	US-10-156-761-1
7	301.4	14.6	9025608	15	US-10-156-761-1

8	280.8	13.6	1479	15	US-10-156-761-6148	Sequence 6148, Ap
9	272.2	13.2	645	20 <td>US-10-716-803-9</td> <td>Sequence 9, Appl1</td>	US-10-716-803-9	Sequence 9, Appl1
10	270	13.1	1362	17 <td>US-10-282-122A-10135</td> <td>Sequence 30135, A</td>	US-10-282-122A-10135	Sequence 30135, A
11	267.2	13.0	2046	22 <td>US-10-501-282-17</td> <td>Sequence 17, Appl1</td>	US-10-501-282-17	Sequence 17, Appl1
12	267.2	13.0	1754382	22 <td>US-10-501-282-17</td> <td>Sequence 6651, Ap</td>	US-10-501-282-17	Sequence 6651, Ap
13	266.8	12.9	1860	9 <td>US-09-738-626-1111</td> <td>Sequence 1111, Ap</td>	US-09-738-626-1111	Sequence 1111, Ap
14	266.8	12.9	2005	21 <td>US-10-494-675-25</td> <td>Sequence 25, Appl1</td>	US-10-494-675-25	Sequence 25, Appl1
15	266.8	12.9	3309400	9 <td>US-09-738-626-1</td> <td>Sequence 1, Appl1</td>	US-09-738-626-1	Sequence 1, Appl1
16	264.8	11.4	42000	19 <td>US-10-159-257A-1</td> <td>Sequence 1, Appl1</td>	US-10-159-257A-1	Sequence 1, Appl1
17	234.8	11.4	63158	16 <td>US-10-292-126A-1</td> <td>Sequence 1, Appl1</td>	US-10-292-126A-1	Sequence 1, Appl1
18	231.6	11.2	1338	17 <td>US-10-282-122A-32104</td> <td>Sequence 32104, A</td>	US-10-282-122A-32104	Sequence 32104, A
19	224.2	10.9	2523	17 <td>US-10-282-122A-17825</td> <td>Sequence 17825, A</td>	US-10-282-122A-17825	Sequence 17825, A
20	220.4	10.7	1257	17 <td>US-10-282-122A-33856</td> <td>Sequence 33856, A</td>	US-10-282-122A-33856	Sequence 33856, A
21	210.2	10.2	2160	14 <td>US-10-138-927-76</td> <td>Sequence 76, Appl1</td>	US-10-138-927-76	Sequence 76, Appl1
22	210.2	10.2	2160	14 <td>US-10-138-927-83</td> <td>Sequence 83, Appl1</td>	US-10-138-927-83	Sequence 83, Appl1
23	210.2	10.2	2160	17 <td>US-10-430-011-76</td> <td>Sequence 76, Appl1</td>	US-10-430-011-76	Sequence 76, Appl1
24	210.2	10.2	2160	17 <td>US-10-430-011-83</td> <td>Sequence 83, Appl1</td>	US-10-430-011-83	Sequence 83, Appl1
25	210.2	10.2	2160	17 <td>US-10-430-011-126</td> <td>Sequence 126, Appl1</td>	US-10-430-011-126	Sequence 126, Appl1
26	201.2	9.8	2159	17 <td>US-10-430-011-122</td> <td>Sequence 122, Appl1</td>	US-10-430-011-122	Sequence 122, Appl1
27	198.2	9.6	1353	17 <td>US-10-282-122A-23903</td> <td>Sequence 23903, A</td>	US-10-282-122A-23903	Sequence 23903, A
28	196.4	9.5	636	15 <td>US-10-156-761-4319</td> <td>Sequence 4319, Ap</td>	US-10-156-761-4319	Sequence 4319, Ap
29	192.8	9.4	225646	19 <td>US-10-470-565-1</td> <td>Sequence 1, Appl1</td>	US-10-470-565-1	Sequence 1, Appl1
30	190.4	9.2	1323	17 <td>US-10-282-122A-19610</td> <td>Sequence 19610, A</td>	US-10-282-122A-19610	Sequence 19610, A
31	189.6	9.2	2190	14 <td>US-10-138-927-75</td> <td>Sequence 75, Appl1</td>	US-10-138-927-75	Sequence 75, Appl1
32	189.6	9.2	2190	17 <td>US-10-430-011-75</td> <td>Sequence 75, Appl1</td>	US-10-430-011-75	Sequence 75, Appl1
33	186	9.0	661	20 <td>US-10-425-115-81629</td> <td>Sequence 81629, A</td>	US-10-425-115-81629	Sequence 81629, A
34	182.2	8.8	2166	17 <td>US-10-430-011-127</td> <td>Sequence 127, Appl1</td>	US-10-430-011-127	Sequence 127, Appl1
35	180.8	8.8	2202	17 <td>US-10-430-011-121</td> <td>Sequence 121, Appl1</td>	US-10-430-011-121	Sequence 121, Appl1
36	178.6	8.7	65140	19 <td>US-10-203-295-1</td> <td>Sequence 1, Appl1</td>	US-10-203-295-1	Sequence 1, Appl1
37	178.6	8.7	125401	19 <td>US-10-203-295-35</td> <td>Sequence 35, Appl1</td>	US-10-203-295-35	Sequence 35, Appl1
38	173.8	8.4	2190	14 <td>US-10-138-927-93</td> <td>Sequence 93, Appl1</td>	US-10-138-927-93	Sequence 93, Appl1
39	173.8	8.4	2190	17 <td>US-10-138-927-93</td> <td>Sequence 93, Appl1</td>	US-10-138-927-93	Sequence 93, Appl1
40	173.6	8.4	1119	15 <td>US-10-156-761-6781</td> <td>Sequence 6781, Ap</td>	US-10-156-761-6781	Sequence 6781, Ap
41	170.6	8.3	2190	14 <td>US-10-138-927-1</td> <td>Sequence 1, Appl1</td>	US-10-138-927-1	Sequence 1, Appl1
42	170.6	8.3	2190	14 <td>US-10-138-927-84</td> <td>Sequence 84, Appl1</td>	US-10-138-927-84	Sequence 84, Appl1
43	170.6	8.3	2190	14 <td>US-10-138-927-85</td> <td>Sequence 85, Appl1</td>	US-10-138-927-85	Sequence 85, Appl1
44	170.6	8.3	2190	14 <td>US-10-138-927-86</td> <td>Sequence 86, Appl1</td>	US-10-138-927-86	Sequence 86, Appl1
45	170.6	8.3	2190	14 <td>US-10-138-927-87</td> <td>Sequence 87, Appl1</td>	US-10-138-927-87	Sequence 87, Appl1

ALIGNMENTS

RESULT 1
US-10-472-587-1
; Sequence 1, Application US/10472587
; Publication No. US20040214274A1
; GENERAL INFORMATION:
; APPLICANT: YANAI, Koji
; APPLICANT: SUMIDA, Naomi
; APPLICANT: MATANABE, Manabu
; APPLICANT: MORIYA, Tateaki
; APPLICANT: MURAKAMI, Takechi
; TITLE OF INVENTION: Transformants Producing Substance P/1022 Derivatives, Methods fo
; TITLE OF INVENTION: Producing The Same And Novel Biosynthesis Genes
; FILE REFERENCE: 2003-1302A/WMC/00144
; CURRENT APPLICATION NUMBER: US/10/472,587
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 82227/2001
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2061
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(2058)
US-10-472-587-1
Query Match 100.0%; Score 2061; DB 20; Length 2061;
Best Local Similarity 100.0%; Pred. No. 0;

Matches 2061: Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	1	ATGCGACGCTTGTGATCGACAACTAGACTCGTTCAACCCACAACCTGTTCAGTACATC	60	
Db	1	ATGCGACGCTTGTGATCGACAACTAGACTCGTTCAACCCACAACCTGTTCAGTACATC	60	
Qy	61	GGGAGGCGCACCGGGCAACCCCGGTGTCGTGCGCCACGACGCGGCTGTGCGGCTG	120	
Db	61	GGGAGGCGCACCGGGCAACCCCGGTGTCGTGCGCCACGACGCGGCTGTGCGGCTG	120	
Qy	121	CCGCTGAGGAGCTTTCGACGCGATCGTGTGTCCCGGGGCGCGGACGCGCGGAA	180	
Db	121	CCGCTGAGGAGCTTTCGACGCGATCGTGTGTCCCGGGGCGCGGACGCGCGGAA	180	
Qy	181	CGGAGCTTCGGAATCAGCCCGCGGCGATCAGCAGCAGCGGCTGTGCGGCTG	240	
Db	181	CGGAGCTTCGGAATCAGCCCGCGGCGATCAGCAGCAGCGGCTGTGCGGCTG	240	
Qy	241	TGCGTCGAGCCACGAGGCGATCGCCAGCTCTTGCGCGAAACGCTCGGCTCGCCCGAA	300	
Db	241	TGCGTCGAGCCACGAGGCGATCGCCAGCTCTTGCGCGAAACGCTCGGCTCGCCCGAA	300	
Qy	301	CCCATGACGCGCGGGCTCCGAGTGCAGGACACCGCGAGGACGTCCTTCGCGGCTGC	360	
Db	301	CCCATGACGCGCGGGCTCCGAGTGCAGGACACCGCGAGGACGTCCTTCGCGGCTGC	360	
Qy	361	CCCTCGCGCTTCAACCGCGGTGCGTACACTCCCTGGCGCGACCGACCTGCCGACGAG	420	
Db	361	CCCTCGCGCTTCAACCGCGGTGCGTACACTCCCTGGCGCGACCGACCTGCCGACGAG	420	
Qy	421	CTCGAACCCCTCGCTGAGACGACGAGGCTGTCATGAGGCTTGCGGACCGCGAAG	480	
Db	421	CTCGAACCCCTCGCTGAGACGACGAGGCTGTCATGAGGCTTGCGGACCGCGAAG	480	
Qy	481	CGCGTGGGGGCGTCCAGTTCCACCGCGAGTCCATCGGACGACATTCGCGCGGAGATC	540	
Db	481	CGCGTGGGGGCGTCCAGTTCCACCGCGAGTCCATCGGACGACATTCGCGCGGAGATC	540	
Qy	541	ATGAGCAACTTTCGCGACCTCGCCCTGCGCCACACCGGAGCAGCGGCGCGAC	600	
Db	541	ATGAGCAACTTTCGCGACCTCGCCCTGCGCCACACCGGAGCAGCGGCGCGAC	600	
Qy	601	TCCCGTACGAACTCCAGTGCAGCGCGGTGACGTCGTGCGGACGCGCAAGAGTACGC	660	
Db	601	TCCCGTACGAACTCCAGTGCAGCGCGGTGACGTCGTGCGGACGCGCAAGAGTACGC	660	
Qy	661	CGGCGCTGCTGCCCGCGGAGGACACACCTTCTGGTGGACAGCACTCCGTCCTGAA	720	
Db	661	CGGCGCTGCTGCCCGCGGAGGACACACCTTCTGGTGGACAGCACTCCGTCCTGAA	720	
Qy	721	GGGCGCTCGGCTTCTCTCTCTCGGAGACGACGCGGCGCGCTCGCGAGTACCTCAC	780	
Db	721	GGGCGCTCGGCTTCTCTCTCTCGGAGACGACGCGGCGCGCTCGCGAGTACCTCAC	780	
Qy	781	TACCGCGTCCGACGCGCGTCTCTCGGTCGCGGCTCCGACGACACAGACCCGAGC	840	
Db	781	TACCGCGTCCGACGCGCGTCTCTCGGTCGCGGCTCCGACGACACAGACCCGAGC	840	
Qy	841	CGGCGCGCTTCTTCACTACCTGAGAGACAGTCCGACGCGAGCGGCTGCCGCTGCC	900	
Db	841	CGGCGCGCTTCTTCACTACCTGAGAGACAGTCCGACGCGAGCGGCTGCCGCTGCC	900	
Qy	901	CCGGAATGCGCTTTCGAGTTCAACTCGGCTACGTCGGCTACCTCGGCTACGAGCTGAG	960	
Db	901	CCGGAATGCGCTTTCGAGTTCAACTCGGCTACGTCGGCTACCTCGGCTACGAGCTGAG	960	
Qy	961	GCGAGACACACCGGCGACCCCGGCGACCGGTCGCCGACCCCGACGCGGCTCTCTTC	1020	
Db	961	GCGAGACACACCGGCGACCCCGGCGACCGGTCGCCGACCCCGACGCGGCTCTCTTC	1020	
Qy	1021	GCGGACCGGCGCATCGCTCTGACACACGAGAGGCTGTGCTACCTGTGCGGCTCTGAC	1080	
Db	1021	GCGGACCGGCGCATCGCTCTGACACACGAGAGGCTGTGCTACCTGTGCGGCTCTGAC	1080	

Qy	1081	CGCGGGGCCACGACGACGCGCGCCGCGCTGCGGAGACGCGCCGACCCCTCAC	1140	
Db	1081	CGCGGGGCCACGACGACGCGCGCCGCGCTGCGGAGACGCGCCGACCCCTCAC	1140	
Qy	1141	GGCTGCGCTGCGCGCCCGCGCGACCCCGCATGTGCTTGGGATTCGCGAG	1200	
Db	1141	GGCTGCGCTGCGCGCCCGCGCGACCCCGCATGTGCTTGGGATTCGCGAG	1200	
Qy	1201	GGGCGCGCGGCTTCGAGCCCGCTGGCGCGCGCGCACGACAGGACGCTACCTCAAG	1260	
Db	1201	GGGCGCGCGGCTTCGAGCCCGCTGGCGCGCGCGCACGACAGGACGCTACCTCAAG	1260	
Qy	1261	CGCATGACGAGTGCCTCAAGAGATCCGCAACGCGAGTCTGACAGATCTTGAC	1320	
Db	1261	CGCATGACGAGTGCCTCAAGAGATCCGCAACGCGAGTCTGACAGATCTTGAC	1320	
Qy	1321	AACATGCTCAACGCGCGACCCGAGCGACGCGCTGCGCTCTACCTCGCGCTGCGCGC	1380	
Db	1321	AACATGCTCAACGCGCGACCCGAGCGACGCGCTGCGCTCTACCTCGCGCTGCGCGC	1380	
Qy	1381	ATCAGCCCGCTCCGTAAGGCGCCCTGCTGAGTTCCCGGAACGTGCGGTGAGCGCC	1440	
Db	1381	ATCAGCCCGCTCCGTAAGGCGCCCTGCTGAGTTCCCGGAACGTGCGGTGAGCGCC	1440	
Qy	1441	TGCGCCGAGCGGTTCTTCAAGATCGGCGCGACGCGCGGTCGAGTCCAAAGCCATCAAG	1500	
Db	1441	TGCGCCGAGCGGTTCTTCAAGATCGGCGCGACGCGCGGTCGAGTCCAAAGCCATCAAG	1500	
Qy	1501	GGGACCGCGCCCGGGGCGGACCCGCGAGAGAGACGACGCGCTCCGCGCGACCTTGCC	1560	
Db	1501	GGGACCGCGCCCGGGGCGGACCCGCGAGAGAGACGACGCGCTCCGCGCGACCTTGCC	1560	
Qy	1561	GGCGGGAGAGAGACCGGGCGAGAACCTGATGCTGACCTGGTCGGAACGACCTC	1620	
Db	1561	GGCGGGAGAGAGACCGGGCGAGAACCTGATGCTGACCTGGTCGGAACGACCTC	1620	
Qy	1621	AACAGCGTCTGCGCGATTCGCTCCGTCAGCTGACCGCGCTCTTGAGGTGAGACTTAC	1680	
Db	1621	AACAGCGTCTGCGCGATTCGCTCCGTCAGCTGACCGCGCTCTTGAGGTGAGACTTAC	1680	
Qy	1681	GGCGCGTGCACAGCTGTGTGACATCCGAGGACGCGCTCGCGCCGACCAAGCAC	1740	
Db	1681	GGCGCGTGCACAGCTGTGTGACATCCGAGGACGCGCTCGCGCCGACCAAGCAC	1740	
Qy	1741	GCGCGCTGCTAGCGCGCGCTTCCCGCGGCTCCATGACCGGGCGCGCAAGAGCGC	1800	
Db	1741	GCGCGCTGCTAGCGCGCGCTTCCCGCGGCTCCATGACCGGGCGCGCAAGAGCGC	1800	
Qy	1801	ACCATGAGATCATGACCGCGCTGAGAGAGCCCGCGGCGCTTACTCCGAGGCGCTC	1860	
Db	1801	ACCATGAGATCATGACCGCGCTGAGAGAGCCCGCGGCGCTTACTCCGAGGCGCTC	1860	
Qy	1861	GGATGCTTGCCTTACGCGCGCGCGCTCAGAGATGTCATCGACACATCGGCTG	1920	
Db	1861	GGATGCTTGCCTTACGCGCGCGCGCTCAGAGATGTCATCGACACATCGGCTG	1920	
Qy	1921	GCGGAGCGGACGAGATTCGCGCGCGCGGCGAGTGTGTCTCTCGACAGAGG	1980	
Db	1921	GCGGAGCGGACGAGATTCGCGCGCGCGGCGAGTGTGTCTCTCGACAGAGG	1980	
Qy	1981	GAGGAGTTTACGAGACCGTGTAAAGCGCGCGCATGTCACGCGCTCGACGCGAC	2040	
Db	1981	GAGGAGTTTACGAGACCGTGTAAAGCGCGCGCATGTCACGCGCTCGACGCGAC	2040	
Qy	2041	GCGGTGCGGCGCGCGGATGA	2061	
Db	2041	GCGGTGCGGCGCGCGGATGA	2061	

RESULT 2
US-10-126-927-68
; Sequence 68, Application US/10126927

Qy	1681	GGGCGCGGTGACACAGCTGTGTTCGACCATCCGGGGAGCGGCTGGCGGCGCCCGGACCCAGACACC	1740
Db	1630	GGCGCCCGTGACACAGCTGTGTTCGACCATCCGGGGAGCGGCTGGCGGCGCCCGGACCCAGACACC	1689
Qy	1741	GCCGCGTTCGACACAGCGCGCGCTTCCCGGCGCGCTCCATGACCGGCGCGCCCAAGAGCGC	1800
Db	1690	GCCGCGTTCGACACAGCGCGCGCTTCCCGGCGCGCTCCATGACCGGCGCGCCCAAGAGAGCGA	1749
Qy	1801	ACCATGAGATCATCGACCGCTGTGAGGAAAGCCCCCGGGGCGTCTACTCCGGGCGCTC	1860
Db	1750	CCCATGAGATCATCGACCGCTGTGAGGAAAGCCCCCGGGGCGTCTTAAACCGGGGCGCTC	1809
Qy	1861	GGATGTGTTCGCGCTTCAGGGGCGCGCGGACCTCAGCATCTGTCATCCGACCATGTGTCTG	1920
Db	1810	GGATGTGTTCGCGCTTCAGGGGCGCGCGGACCTCAGCATCTGTCATCCGACCATGTGTCTG	1869
Qy	1921	GCCGACGCGCACAGGGGAGGATTCGGGCTGGGGCGGGGCGATCTGTGTCCCTCCGACCAAGAG	1980
Db	1870	GCCGACGCGCACAGGATTCGGGCTGGGGCGGGGCGATCTGTCCCTTCGACCAAGAG	1929
Qy	1981	GAGGAGTTCACCGAGACCGTGTAAAGGCCCGGCGCATGTGTCAACCGGCTTCGACGGGAGC	2040
Db	1930	GAGGAGTTCACCGAGACCGTGTGTCAAGGCCCGGCGCATGTGTCAACCGGCTTCGACGGGAGC	1989
Qy	2041	GCCGTGCGGGCGCCCGCATGA	2061
Db	1990	GCAGTGGCGGGCGCCCGCATGA	2010

RESULT 3
US-10-12

US-10-126-927-67

Sequence 67, Application US/10126927

Publication No. US20030082575A1

; GENERAL INFORMATION:

APPLICANT: The Scripps Research Institute

APPLICANT: Schultz,

APPLICANT: Wang, Lei

APPLICANT: Anderson, John C

APPLICANT: Iain David B

APPLICANT: MacLary, Thomas

APPLICANT: Meggers, Eric L

APPLICANT: Mehl, Ryan A

APPLICANT: Pasternak, Miro

APPLICANT: Santoro, Stephen

APPLICANT: Zhang, Zhiwen

FILE REFERENCE: 54-000130U5

CURRENT APPLICATION NUMBER: US/10/126.927

CURRENT FILING DATE: 2002-04-19

PRIOR APPLICATION NUMBER: US 60/285,030

PRIOR FILING DATE: 2001-04-19

PRIOR APPLICATION NUMBER: US 60/355,514

NUMBER OF SEC ID NOS: 70
PRIOR FILING DATE: 2002-
;

```

; NUMBER OF SEQ ID NOS: 13
SOEHWAPF: Patent version 3.1

```

DOJ ID NO 67

LENGTH: 12391

TYPE: DNA

ORGANISM: Plasmid pSC101, Streptomyces venezuelae pAPABC

US-10-126-927-67

Query Match 86.0%; Score 1772.2; DB 14; Length 12391;

OY 1 ATGGCGACGCTTCTGATGCAACAATACGACCTGGTACACCAACACTGTTCCAGTACATC 60
 Db 24 ATGGCGACGCTTCTGATGCAACAATACGACCTGGTACACCAACACTGTTCCAGTACATC 83
 OY 61 GGCGAGGCGCACCGGGCAACCCCCCGTGTGATGTCGCCAAGACGCGCATGTTGCGGAGCTG 120
 Db 84 GGCGAGGCGCACCGGGCGAGCCCCCG----TCGTGCGCAACGAGCGCGCATGTTGCGGAGCTG 140

QY	121	CCCGTCGAGGACCTTCGACAGCGATCGTGTCTCCCGGGCCCCGGGACGCGCGGAA	180
Db	141	CCCTCTGAGGACTTCGACGCGATCGTGTCTCCCGGGCCCCGGGACGCGCGGAA	200
QY	181	CGGACCTTCGGAATCAGCCGCGGGGCGATTCACCGACAGCGGCTGCCGCTCGGCGTC	240
Db	201	CGGACCTTCGGAATCAGCCGCGGGGCGATTCACCGACAGCGGCTGCCGCTCGGCGTC	260
QY	241	TGCGCTCGGCGCACGAGGGCATGCGCCAGCTCTTCGCGGAAACCGTCGGCTCGCCCCGAA	300
Db	261	TGCGCTCGGCGCACGAGGGCATGCGCCAGCTCTTCGCGGAAACCGTCGGCTCGCCCCGAA	320
QY	301	CCCATGACGAGCGCGGGGTCTCCGAGGTGGGGACAACCGCGAGGACGTCTTCGCGGGAGCTC	360
Db	320	CCCATGACGAGCGCGGGGTCTCCGAGGTGGGGACAACCGCGAGGAGTCTTCGCGGGAGCTC	380
QY	361	CCCTTCGCGGTTCAACCGCGGTGCGCTACCACTCCGTGCGCGCAACGACTTCGCCGAGAG	420
Db	380	CCCTTCGCGGTTCAACCGCGGTGCGCTACCACTCCGTGCGCGCAACGACTTCGCCGAGAG	440
QY	421	CTTCGAACCCCTCGCTGCGAGCGAGACGAGGGGTCTCATGCGCTTCGCGCACCGGAGAA	480
Db	440	CTTCGAACCCCTCGCTGCGAGCGAGACGAGGGGTCTCATGCGCTTCGCGCACCGGAGAA	500
QY	481	CGCGCTGAGGGGAGTTCAGAGTTCACCCGAGTTCATTCGCGACGCACTTCGCGCGGAGATC	540
Db	500	CGCGCTGAGGGGAGTTCAGAGTTCACCCGAGTTCATTCGCGACGCACTTCGCGCGGAGATC	560
QY	541	ATGCGCAACTTCGCGACCTTGCGCTCGCCCAACCGCGGAGACGCGCGCACGCGGCGGAC	600
Db	560	ATGCGCAACTTCGCGACCTTGCGCTCGCCCAACCGCGGAGACGCGCGCACGCGGCGGAC	620
QY	601	TCCCGGTGCAACTTCAGTGCAGCGCGGTGCAACGCTGCGCGGAGCGCGGAGAGGTACGC	660
Db	620	TCCCGGTGCAACTTCAGTGCAGCGCGGTGCAACGCTGCGCGGAGCGCGGAGAGGTACGC	680
QY	661	CGCGCTGCGCGCGCGCGGAGGAGACCACTTCGTGCTGACAGCAAGCTCCGCTCGAA	720
Db	680	CGCGCTGCGCGCGCGCGGAGGAGACCACTTCGTGCTGACAGCAAGCTCCGCTCGAA	740
QY	721	GCGCGCTCGCGCTTCCTTCCTCGGAGACGACCGCGGCTCGCGATTAAGTCAAC	780
Db	740	GCGCGCTCGCGCTTCCTTCCTCGGAGACGACCGCGGCTCGCGATTAAGTCAAC	800
QY	781	TACCGCGTGCAGACGCGGCTGCTCGTTCGCGGAGTTCGACGCGACCAACGACCGGAGC	840
Db	800	TACCGCGTGCAGACGCGGCTGCTCGTTCGCGGAGTTCGACGCGACCAACGACCGGAGC	860
QY	841	CGGCGCCCTTCCTCAACTACCTGAGAGAGCAAGTTCGACGCGGACGCGGCTCCGCTCGCC	900
Db	860	CGGCGCCCTTCCTCAACTACCTGAGAGAGCAAGTTCGACGCGGACGCGGCTCCGCTCGCC	920
QY	901	CCCGAACCGCGCTTCGAGTTCAACTCGTGGCTACGCTCGGCTACCTCGGTTAGAGGTGAAG	960
Db	920	CCCGAACCGCGCTTCGAGTTCAACTCGTGGCTACGCTCGGCTACCTCGGTTAGAGGTGAAG	980
QY	961	GCGAGACCAACGCGGACCCCGCGCACCGGATCCCGGACCCCGACCGCGGCTTCCTTC	1020
Db	980	GCGAGACCAACGCGGACCCCGCGCACCGGATCCCGGACCCCGACCGCGGCTTCCTTC	1040
QY	1021	GCGGACCGCGGCTTCGCTTCGACCAACGAGAAAGCTGCTGATCTGCGGCTTCGAC	1080
Db	1040	GCGGACCGCGGCTTCGCTTCGACCAACGAGAAAGCTGCTGATCTGCGGCTTCGAC	1100
QY	1081	CGCGCGGCGCACGACGAGCGCGCGCTGCGTGCAGGAGAGGCGCGAGACCTTCAC	1140
Db	1100	CGCGCGGCGCACGACGAGCGCGCGCTGCGTGCAGGAGAGGCGCGAGACCTTCAC	1160
QY	1141	GAGCTGAGCGTTCGCGCGCCCGGCGAGCGACCCCGGCAATGCTTCGCGATCCCGGAG	1200
Db	1160	GAGCTGAGCGTTCGCGCGCCCGGCGAGCGACCCCGGCAATGCTTCGCGATCCCGGAG	1220

QY	775	CTGACCTTACCGCGCTTCGCGGACAGCGGCTGTCTCCGTCCGCGGCTTCGACCGCACACGACC	834
Db	19450	GCCAAAGGCCGACGTGACGCGCCGGAAAGGTACAGTGAAG--CCGACGCGCGCACACAC	19506
QY	835	CGGACGCGCGCGCCCTCTTCAACTACCTGAGAGAGAGCTCCAAAGCCGACGCGGTCGCC	894
Db	19507	ACGGTCCGAGAGCGCTTCTCGACCTGGCTGAGAGACGACTGGCGGGGGCTGG--CACC	19563
QY	895	GTCGCCCCCGAACCTGCCCCCTTCAGGTTCAACTTGGGCTACGTGGGTACTCGGCTACCGAG	954
Db	19564	GAGGTGCCCCGAACTTCCCTTCGCGTTGCGCCTCGGCTGGGTTCGGCTGCTCGGCTTACGAG	19623
QY	955	CTGAAAGCGGAGACCAACCGGCGACCCCGCGCACCGGTCGCCGACCCGACCGCGCGCTTC	1014
Db	19624	TTGAAAGCCGAGTCCGACGCGCGACGCGCGCACCGCTCAACCGATCCGACGCCCTGCTG	19683
QY	1015	CTCTTCCCGCACCGCGCCATCGCCTTCGACCAACAGAAAGCTCTGCTACTCTTGCC	1074
Db	19684	GTCTTCCCGCACCGGCGCTGTGTCTGAGCAACCGACCCGCGACACTTACTCTGCTGGCG	19743
QY	1075	CTCGACCGCGCGGGCCACGACGACGCGCCCGGCTGGCTGGGGAGACGCGCCGAGACC	1134
Db	19744	CTGTGTGAGAGACGACGCGCGGAGCCGAGGACGCGGCTGTGCGCGCGGCTTCGCGCAC	19803
QY	1135	CTCAACCGGCTGCGCGCTCGCGGCCCGCGGACCGCACCCCGCCCACTGCTTCCGAGATC	1194
Db	19804	CTGGAAGGCATTCGCGGGGCGGAGGC-----CGAGCGCTGC	19839
QY	1195	CCCGAGCGCGCGCGCGCTTCGCGCCCCCTTGGCCCGCGCGCGCCACGACGACGCTTAC	1254
Db	19840	CCCGAGCGCGCGCGCTGTCACGACGGGTCCGATGAGCTGTGCGCACACACCGGACGCTTAC	19859
QY	1255	CTCAAGCGCATTCGACGAGTGCTCTCAAGAGATCCGACGCGCAGTCTGATCGAGATCTAC	1314
Db	19900	CTGAAGCTGATCGACGCTCTGCGACGAGAGATGACCGCGGGAGACCTACGAGGCTTCC	19958
QY	1315	CTGACCAACATGATCACCAGCGCGCGACCGAGGCGACCGCCCTGCAGCTTACTCCGCGCTG	1374
Db	19960	CTGACCAACATGCGCGGAGCGGACCAACGACCTTCAACCCCGGCGCGGCTTACCGCGCGCTG	20019
QY	1375	CGCGCCATCAGCGCGCGCTTCCGTAACGGGCGCCCTGTGAGTTCGCCGAACTGTCCGCTG	1434
Db	20020	CGCGCGGAGACCCCGCGCGCTTCCGCGCTTCTGTGACTTTCGAGCCCATGCGCGGCTC	20079
QY	1435	AGCGCTTCGCGCGGAGCTTCTTCAACGATCGGCGCGACCGCGCGCTGCGAGTCCAAAGCC	1494
Db	20080	AGCAGCTTCGCGAGCGGTTCTGCGCATCGACCGGACGGGCGGATGAGATCAAAGCCG	20139
QY	1495	ATCAAGGGGACCCGCCCCCGGAGTGGGACCGCGGAGGAGACGACGGGCTCCGCGCCGAC	1554
Db	20140	ATCAAGGGGACGCGGCGCACGCGGCGCGCACCCCGAGAGAGACCGCGCTCTGTAACGCGC	20199
QY	1555	CTGCGCGGCGCGGAGAGAGACCGGCGCGGAGACCTGTATGTCGTGACCTTGGTCCGCAAC	1614
Db	20200	CTGGCTCACCTGCGAAGAGACCGGCGCGGAGACTGATGATCTGTCACTTGGTCCGCGAC	20258
QY	1615	GACCTCAACAGCGCTCTGCGCGATCGGCTCCGTCAACGCTGCCCGGCTCTTCAAGTGAAG	1674
Db	20260	GACCTGGGCGGCTGCGCGGAGTGGGCTCGGTGCTGCGGACCCGAGTTCAGAGTCCAG	20319
QY	1675	ACCTTACGCGCCCGTGTGACCAAGCTGTGTGACCAATCCGGGAAAGGCTGTGGGCCCGGAC	1734
Db	20320	ACGTACGCGACCTGTGACCAAGCTGTGTGATGATCCGTCAACGCGCGGCTGCGCGGACGAC	20379
QY	1735	AGCACCGCGCGCTCTGATCGGCGCGCTTCCCGCGCGCTTCACTGACCGGCGCGCCAG	1794
Db	20380	AGCCCGTGGGCGGCTTCGGGCGGCTTCCCGCGCGATCTGATGACCGGCGCGCGGAAG	20439
QY	1795	AAGCGCACATGAGATCATGACGCGCTTGAAGAGAGCCCGCGGCGCTTACTTCGCG	1854
Db	20440	ATTCGCGACATGAGATCATGACCGGCTGGAAGAGCGGCGCGCGGCTGTCTTCTCGGCG	20499

Qy	1855	GCGCTGGAGTGTGCGCCCTCAGCGGCGCGCGACCTCAAGAAATCGTCATCCGCACATC	191
Db	20500	GCCATCGGACTACTTCTCCCTCACCGSGCGGTGAAGCTTGTCATCGATGCCAAGCGTG	2055
Qy	1915	GTGCTGGCGCGACCGGCGGAGTTGCGGCGTGGCGGCGGCGCATCGTCTCTCCGAC	1974
Db	20560	GTGCTCAGCGCGCGGCGGAGGCTGCGGTACGGGGTGGCGGCGCGCGTCAATCGGCTCTCCGAC	2061
Qy	1975	CAGAGAGAGATTCAACCGAGACCGTGTGAAGGCCCGCGCATGTCAACGCCCTTGAC	2034
Db	20620	CCGGCCCAACGAGTTTCGAGAGAGACGCGCGGTCAAGGCCCGCCCGCGTGTGCGTCTCTCGAC	2067
Qy	2035	GGCAGCGCGCTGG	2047
Db	20680	ACCgccttcccg	20692
 RESULT 5 US-10-156-761-1167 ; Sequence 1167, Application US/10156761 ; Publication No. US20030119018A1 ; GENERAL INFORMATION: ; APPLICANT: OMURA, SATOSHI ; APPLICANT: IKEDA, HARUO ; APPLICANT: ISHIKAWA, JUN ; APPLICANT: HORIKAWA, HIROSHI ; APPLICANT: SHIBA, TADAYOSHI ; APPLICANT: SAKAKI, YOSHIYUKI ; APPLICANT: HATTORI, MASAHIRA ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES ; FILE REFERENCE: 249-262 ; CURRENT APPLICATION NUMBER: US/10/156, 761 ; CURRENT FILING DATE: 2002-05-29 ; PRIOR APPLICATION NUMBER: JP 2001-204089 ; PRIOR FILING DATE: 2001-05-30 ; PRIOR APPLICATION NUMBER: JP 2001-272697 ; PRIOR FILING DATE: 2001-08-02 ; NUMBER OF SEQ ID NOS: 15109 ; SEQ ID NO 1167 ; LENGTH: 2196 ; TYPE: DNA ; ORGANISM: Streptomyces avermectilis ; FEATURE: ; NAME/KEY: CDS ; LOCATION: (1) ..(2196) US-10-156-761-1167			
 Query Match 29.4%; Score 606.2; DB 15; Length 2196; Best Local Similarity 58.3%; Pred. No. 6.6e-123; Matches 1262; Conservative 0; Mismatches 763; Indels 138; Gaps 6;			
Qy	2	TGGCGAGCCTTCTATCGAACATAAGACTGTTCAACCCACAACCTGTCAGTAATCG	61
Db	2	TGAAGACCTCTCTATGACATTAACACTGTGACACTGACACTTTCAGTGAATCG	61
Qy	62	GCGAGGCCACCGGCGCAACCCCCGTCGTGTGCCCCAAGACGCCGACTGTGCGGCTGC	121
Db	62	CCGAGGTGAACGGGAGAGAGCCGGTGTGTCCTCAATGACGCCCGCGGACGACATTC	121
Qy	122	CCG---TCGAGGACTTGCAGCGCATGTCGTGTCCTCCCGGCGCCCGGACGCCGACCGG	178
Db	122	CGGATCTCCGGGATTTGCGCAAACGTGTGTGTGTGTCGGGGGACCGGACACACCCGCGAAGA	181
Qy	179	AACGGGACTTCGGAATCAGCCGCGGCGCGATCAACCGACAGCGGCTGCGGCTCTCGGCG	238
Db	182	GACGTGATTCGGGATGTCGTGGCGAGCTGCTCGCCGCGGCCCGCTGCTGCTGGGCG	241
Qy	239	TCTGCTCTGCGCAACAGGCGCATCGCCAGCTCTTTCGCGGAAACGTCGAGCTCGCCCGG	298
Db	242	TCTGCTCTGCGCAACAGGCGCATTCGCGCGCGGAGAGCGGCGCTGGGTGAACCCCGGCAACGG	301
Qy	299	AACCCATGCAAGCGCGGAGTCTTCGAGGTGCGGCAACAGCGGAGAGAGTCTTTCGCGGCGC	358

Db	302	AGCCCCGACACGGGCAATCTGTCCACCGCTCCGGACAAAGGCAAGAACCTGTTCGAGGGCC	361
OY	359	TCCTCTCGCGCTTTCAACCGCGCTGTCGCTAACACTCTCTTGCGGCACACGA---CTCTCCCG	415
Db	362	TGCCCCAGACGATTCAACCGCGCTCCGCTACCACTCTCTGCTGTAAGGGAACCGCTGCGCAC	421
OY	416	ACGAGCTCGAACCCCTCGCTGGAGGAGACGAGGGGTGTCTATGGGCTCTGGCGACCCGG	475
Db	422	TGAGCTTGAGCCCAACCGCTTGTGGGAGGAGACGGGTGTGATGGGATCTGCGGACACCGCT	481
OY	476	AGAAACCCGCTGTGGGCGCTCCAGTTTCAACCCGAGTTCATTCGACAGCATTTCCGCGGG	535
Db	482	CCGGCGCGCTGTGGGGGGGTGAGTTTCAACCGGAGTCCGTCTCAACCGATTCCGCGCAC	541
OY	536	AGATCATGGCCAACTTTCCGCGACCTTCGCTTCGCCCAACCACTGGG-----	580
Db	542	GGATGTTGTGAATTTCCGCGAGCTGACGGCTTCCCGCGCCGCGAGACTCGACGGACA	601
OY	581	-----	580
Db	602	AGGCCCTGACGTTCGGAACGCGCACGACCGACCCCGCTGCGGCCCTGTTCCCAACG	661
OY	581	-----CAAGGCGCACGAGGCGCACTCC	604
Db	662	CCGAGGGGGCCGAGGACAGACCCCGCGTGTGTGTCCCGCGGCCCTGTGCTGTGCGCCCGT	721
OY	605	CGTACGAATCTCCAGTGTGCGCGCGCTGACAGTGTCTGCGGACCGCGAGAGGTACGCGCG	664
Db	722	CCCAACGCTTTCACACCGCGCGGATTCGAGTGGCGTGTGACGCGAGGCGGCTTCAACC	781
OY	665	GCTGCTGTCCCGGCGAGGGCACACAGTTTGTGCTGACAGCAGCTCCGTCTTCGAAGGCG	724
Db	782	GGATGTACACCGACCGCGCCCGCGGTTCTGTGCTGACAGCTCCCGGAGTCGAGGAGGAC	841
OY	725	CCTCGCGCTTCTCTTCTCTGGCGACGACCGCGCGCTGCGCGAGTACCTACCTAAC	784
Db	842	AGTCCCGTTCTCTGTTCTTCCGTGACGGACCGGACCGCTGCGCAGATTCTGATCGATTACG	901
OY	785	GCGTGCAGACGGCGGTGTCTTCCGTCCGCGCGCTTCGACGCGACCAACGACCCGAGCGCGAC	844
Db	902	ACGTGAGAGCGGCTCTGTGAGATTCGAGCGGCGCGGGCGGCGGTGGCAAGTCAAGG	961
OY	845	GCCCCCTTTTCACTACTCTGAGAGACAGCTGGAACCGCGACGGGTCCCGTCCGCCCG	904
Db	962	CGAGGTTTGCACCTACCTGAACCGGACGAGTTGTGTACCCGTCAGGT---GGATGTACGCG	1018
OY	905	AACGTCCCTTGAAGTTCAACCTCGGCTAGTCCGCTACTCTGGGCTACGAGCTGAAGCGCG	964
Db	1019	GACTGCGCTTGCATTCACCGCGGGGTATGTGGGCTACTTGGGCTACGAGGTAAGAGGCG	1078
OY	965	AGACCACCGGCGACCCCGCGCACCGGTCCCGCACCCCGACGCGCGCTTCTCTTCCGCG	1024
Db	1079	ACTGTGAGTTCCGTCAACGCTGATTCGGGCGCGAACCTCCGACGCTGTGCTGTTCGCG	1138
OY	1025	ACCGCGCATGTGCTCTGACCAACGAGAAAGCTGTCTTACTCTGTGCGCTCTGACCGCC	1084
Db	1139	ACCGGCTGATGCAAGTGAACATCACTGACGAGA-----GCCACTACGCG	1183
OY	1085	GGGGCGAAGAGACGCGCGCGCGCGCGCTGTGCGGAGAACGCGCGAGACCTTCAACCGGC	1144
Db	1184	TCTGTCTTGCAGGAACAACCGCGAGGGCGCACAGCAAGCGGCGCACTGTGCTCAACGCG	1243
OY	1145	TGAGCGTCCGCGCCCGCGCGACGACCCCGCATAGTCTTTCGGAGATCCCGAGGCGG	1204
Db	1244	CGATGCGCCACACTCAATTTCGTGTCTTGGCGCAACCGACCGCCCGCGCGCGACGCG	1303
OY	1205	CGGCGGCTTCCGCGCCCTGTG---CCGCGCGCGCACGACAGAGACGCTTAACTCAAGC	1261
Db	1304	CACCCCGCTCGATCCCGTGGAGCGGTGATGTGTCGCGACGATCACTAACCTCCGCG	1363
OY	1262	GCATGAGAGGTGCTCAAGAGATCCGCAACCGGGAATCGTACGAGATTTGCTGTGACCA	1321
Db	1364	ACATGGGACCTGCAAGCGGGAACTCAAGACGGGACCAAGTTTACGAGATCTGTCTGACCA	1423

QY	1322	ACATGTCACGCGCCGACCGAGCGGCGGACCGGCTTCGCTTACCTCCGCGTGGCGCA	1381
Db	1424	ACGCGCCAGGCTGCCCCCCCCCGTACGACCCCTTACGACTTTCACCGGTACTGGTGTCT	1483
QY	1382	TCAGCCCCGTCCCGTACGCGCGCCCTGTGAGTTTCCCGAACTGTGGTGTGAGCGCTT	1441
Db	1484	TCGACCCCGGCCCGTACCGCGGCTTATCTGCGGTTGCGTGACTTGCACTTGCCCGGCTGT	1543
QY	1442	CGCCCCGAGGTTTCCCTACGATTCGGCCCGGACGGCGGCTGAGTTCGAAGCCCTAACG	1501
Db	1544	CCCCGAGCGTTTCTTCTGCGATCACCGCGAGCGGCTGCGAGGCGCCGAGCCCGTCAAG	1603
QY	1502	GGAACCGCGCCCCCGGGGCGGACACCGCGAGAGAGAGACGAGCGGCTCCGCGCCGACTGGCCG	1561
Db	1604	GCAACCGCGCCCCCGCGCGAGCGCGCGGAGAGAGAGAGCCCGGCTACCGGACCGCTCACCA	1663
QY	1562	GCCGGAGAGAGAACCGGGCCGAGAACCTGATGATCGTTCGACTTGTCGACGACCTCA	1621
Db	1664	CGSACGACAAAGACCGCGCGGAGAACCTGGTGAATGCTGACCTGCTCGCAACGACCTGG	1723
QY	1622	ACAGGCTTGGGCGCATCGGCTCCGTCCACGTCGCCCCGCTCTTGAGGTGAGACCTACG	1681
Db	1724	GCCGGCTTGGCGGACCGGAGCGGTGAAGTCAACCGCTGATGAGCGACCGAGCGTTCG	1783
QY	1682	CGCCCGTSCACGAGCTGGTTCGACCATCCGGGGACGGGCTGCGGCGCGGACCGACGACCG	1741
Db	1784	CCACCGTGCACGATGTTGTCTTCACCGTCAAGGGGCGGCTGCGAGGGGCACTCGGAGCG	1843
QY	1742	CCGCGCTGCGTACGCGCGCGCTTCCCGGCGGCTTCATGACCGCGCGGCCCAAGAGCGCA	1801
Db	1844	TGACCTGGTACGATGCTGCTTCCCGGAGGTTGAGTGAACGGGGCCCGAATCTGGCA	1903
QY	1802	CCATGAGATTCATGACCGGCTGGAGAGAGGCCCCCGGGGCGTATCTTCGGGGCGCTCG	1861
Db	1904	CGATGGAATCATGACTGCTGTGAGACCGAGGCCCGGGGGTGTACTCCGGAGCCATCG	1963
QY	1862	GATGATTCGACCTCAGCGGCGCGCGCGACTCAGACATGCTCATCGGACCATCGTGTGCG	1921
Db	1964	GCTACTTCGGGTGAGTGGGGGCGCGGACTCTGCGCATGCGAGCGGACCGGCGGTTC	2023
QY	1922	CCGACGCGCAGCGGAGTTCGGCGCTCGGCGGGGCGATGCTGCCCTTCGCGACGAGAGG	1981
Db	2024	CCGACGGGAGATGCACTCGGCGCGGCGGATGCGATGCTCTCGGCTCCGATCCGGTGG	2083
QY	1982	AGGAGTTCACCGGACCGTGTAAAGCCCGCGCATGTCACCGGCTTCGACGCGGACG	2041
Db	2084	GCGAGTACGAGAGATGCTGTGAAGAGGCGCGCACGATGCGGCGCCACCGCGACCGGA	2143
QY	2042	CGG 2044	
Db	2144	TCG 2146	

```

: PRIORITIZING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 1
: LENGTH: 9025608
: TYPE: DNA
: ORGANISM: Streptomyces avermectilis
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (4187715)
: OTHER INFORMATION: a, t, c, g, other or unknown
: US-10-156-761-1

```

Query Match	29.4%	Score 606.2;	DB 15;	Length 9025608;
Best Local Similarity	58.3%	Pred. No. 1.8e-123;		
Matches 1262; Conservative	0;	Mismatches 763;	Indels 138;	Gaps 6;

OY	2	GGCGCAGCTTCTGATGCGAACAACTAAGACTGGTTACACCAACACTGTTCCAGTAACATCG	61
Db	1470396	TGAAGACCTCTCTCATGCACATTAACGACTGTACAGTAACACTGTTCCAGTTATCG	1470357
OY	62	GCGAGGCCACCGGCGAACCCCGCTGCTGTGTCGCAAGACGCCGCACTGTGTGCGGTGC	121
Db	1470536	CCGAGGTGAACGGGCGAGGAGCGCGGTGGTGGTCTTCAATGAGCCCGCGGAGCACTTC	1470477
OY	122	CCG---TCGAGGACTTGCAGCCGATTCGTCTGTGTCCCGGGGCTCCGCGAAGCCCGACCGG	178
Db	1470476	CGGATCTCCGGGATTTTCGCCAACGTGTGTGTGTGCGCGGGGCTCGGACACCTCCGGAAGA	1470417
OY	179	AACGGGACCTTGGGATCAGCGCGCGGGGCGATCACCGACAGGGGCTCGCTCTCGGCG	238
Db	1470416	GACGTGACTTGGGCATCGTCGGGCGAGTGTCTGGCGCGGCCCCCGTGTCTGTGGGGC	1470357
OY	239	TCTGCTTCGCGCACCGAGGCATCGCCCACTCTTCGCGGAAACCTGTGCTGCGCCCG	298
Db	1470356	TCTGCTTCGCGCACCGAGGCATTCGCGCGGGGAGCGGCTGTGGTGAACCCCGCACCGG	1470297
OY	299	AACCATGACAGGCGGGTCTCCGAGGTGGGGCACACCGCGAGAGAGTCTTCGGGGGGC	358
Db	1470296	AGCCCGCGACAGGCATTTGTTCACCGTCCGGCAACAGCGCAAGAACCTGTTCAAGGGCC	1470237
OY	359	TCCCTCGCGCTTACCGCGCGTGGCTAACACTCCCTGCGCGCACGAA---CCTCCCG	415
Db	1470236	TGCCCGACAGTTACCGCGGTCCGCTACCACTCCGTGGGTACGGGAACCGCTGCAC	1470177
OY	416	ACGAGCTCGAACCCCTGCTGCTGAGAGCGACGAGGCTGTATATGGCTCTGGGCAACCGCG	475
Db	1470176	TGAGCTGTGAGACCCACCGCTGTGGCGGAGAGACCGCGTGTATATGAGACTGTGGGCAACGCT	1470117
OY	476	AGAAGCGCCTTGGGGCGCTGCAGTTCCACCCGAGTGCATGCGAGAGACTTCGCGCGGG	535
Db	1470116	CCCGGCGCGTGTGGGGGTGAGATTCCACCCGGAGTCCGTCTACCGAGTTGGGCAAC	1470057
OY	536	AGATCATGGCCAACTTCGCGCACTTCGCGCTTGCCACCAACCGG3-----	580
Db	1470056	GGATGTTGTAACATTCCGCGAGCTGACCGGCTCCCGGCGCGGAGACTCGACGGGCA	1469957
OY	581	-----	580
Db	1469996	AGGCCCGTAGCTCCGGAACGCGCACAGACGAGCCCGCTGCGGCCCCCGTCCCAAGCG	1469937
OY	581	-----CAGCGCGCACGGGCGGAGTCC	604
Db	1469936	CCGAGGGGGCGGGGAGACCCGCGGTGCTGTGCTCCCGGCCCCGCTGTGGCCCGT	1469877
OY	605	CGTACGAACCTCACGTGCGCGCGTTCGACGTGTGCGCGAGACGCGGAAGAGTACGCGCG	664
Db	1469876	CCCAACCGCTTCAACACCGGGGAGATTCGACGTGGCCGTGAGACGCGGAGCGGCTTTCACC	1469817
OY	665	GCTGCTGCTCGGCGAGGGGACACAGTTCTGAGCTGAGACAGAGCTCCGCTCTGAAAGGCG	724
Db	1469816	GGATGTACACCGAGCGCGCCCGCGGCTTCTGTGAGTACAGCTCCCGGAGTGAAGAGGGAC	1469757
OY	725	CCTGCGGCTTCTCTCTTCGAGCAAGACGCGGCGCGCTGCGCGAGTACTTCACTTACC	784

Db	1469756	AGTCCCGTTTCTCGTTCTTGGGTGACGGCACCCGAGCGCTGGCCGAGTTGCTACGGTACG	1469697
Qy	785	GCCTGCGCGAAGCGGCGTGTCTCCGTCCGCGGCTTCGACGGACACACGACCCGAGCGCGC	844
Db	1469696	ACGTGGAAGAGCGGCTCTGTGAGATCGAGCGGCGCCGGCGCCGTCGTGCAAGTCAAGG	1469637
Qy	845	GCCCCCTTCTCACTTACTTGGAGAGCACTCGAACGCCGACGGGTCCCGTGGCCCCG	904
Db	1469636	CCACCGTCTTCTGACTTACTTAAACGGGCAATTGGTGAACCCGTCAAGT---GGATGCCACGG	1469580
Qy	905	AACGCGCTTTCGATTGACCTCGGCTACGTCGCGCTACCTCGGCTACGAGCTTAAGGCGG	964
Db	1469579	GACTGCGGTTTGACTTCAACGGCGGGTATGTGGGCTACTTGGCTAGAGGTAAAGCCG	1469520
Qy	965	AGACCAACGGGAGACCCCGCGACCGGTCGCCCGACCCCGACCGCGCGGTCCTTCGCGG	1024
Db	1469519	ACTGCGGATTCGTGTAACCGTGTATCGGGCCGGAATCCCGACGCGCTGTGGCTTTGGCG	1469460
Qy	1025	ACCGCGCCATTCGCTTGCATCCACCAAGGAGCTGTGCTACTGTGGCTTCGACCGCC	1084
Db	1469459	ACCGGCTGATTCGAGTGAACCATCAGACGCGA-----GCCACCTTACCGCC	1469415
Qy	1085	GGGCGCAAGACGAGGCGCCCGCGCTGTGGGAGAGCGCGCGAGAACCTTCACCGGCC	1144
Db	1469414	TCTGTCTGCGCGAACAACCCGCGAGGCGCCACAGCAAGCGCGACTGGCTGACGCGG	1469355
Qy	1145	TGGCGCTTCGCGCGCCCGGCGAGCGCACCCCGCATGTCTTCGGGATCCCGAGCGG	1204
Db	1469354	CGATGCGCCGACTCAGTTTGTGTCTCTGGCCAAAGCGACCGCGCCCGCGCGGCGACCG	1469295
Qy	1205	CGCGCGCTTCGCGCCCCCTGG---CCGCGCGCGCCACGACMAAGACGCTTACTTCAAGC	1261
Db	1469294	CACCCCGCTTCGATCGCGTGAAGCGCGTGGCTGTGTCGCGACCGTACGACTTACTTCGCGG	1469235
Qy	1262	GCATCGACGAGTGTCTCAAGAGATTCGGCAACGCGAGTGTGTGAGATCTTGCTGACCA	1321
Db	1469234	ACATCGGGGACTCTGMAAGCGGGAATCAAGAGCGGACCACTTACGAGATGTGCTGACCA	1469175
Qy	1322	ACATGTGTCAACGCGCGACGAGCGGCGCGGCTTCGCGCTTACTCGCGCTCGCGGCA	1381
Db	1469174	ACGCGGCGAAGCTGTCCCGCCCGCCGTAAGACCCCTTACGACTTTCACCGGTTACTGCTGTC	1469115
Qy	1382	TCAACCCCTTCGCGTACGAGCGGCGCTGTCTGATGTCCCGAATGTGTGCTGAGCGCT	1441
Db	1469114	TGCAACCGGCGCGCTGACGCGGCTTACTGTGGGTTGCGTGTGACTGATGTGGCGGCTGT	1469055
Qy	1442	CGCCCGAAGGTTTCTTCAAGTTCGAGCGGCGGCGGCGGCTGTGAGTCCAGCCCATCAAG	1501
Db	1469054	CCCCGAGCGTTTCTCGGATTCACCGCGACCGCGCTGTCCGAGCGCACGACCCGTCAAG	1468995
Qy	1502	GGACCGCGCCCCCGGGGGCGGACCGCGGAGAGGAGCGGCTTCGCGCGCGACCTGGCGG	1561
Db	1468994	GCATCGCGCCCCCGCGGAGAGCGCGGAGGAGGAGCGCCCGCTTACGAGAGCGGCTTACCA	1468935
Qy	1562	GC CGGAGAGAGACCGGAGCGGAGACTGTATGTCTGTGACTGTGTCCGACGACTCA	1621
Db	1468934	CGGACGACAGAACCCGCGCGGAGAACCTGTGTATGTGTGACTGTGTCCGACGACTTGG	1468875
Qy	1622	ACACGCTTGTGCGCATTTGGCTTCGTCAAGTGTGCCCGGCTTTCGAGGTGGAACCTTACG	1681
Db	1468874	GCCGCGTCTGCGGACCGGAGCGGTGAAGTTCACCCGCTGATGGCGACGGAACGTGCG	1468815
Qy	1682	CGCCCGGACCAACTGTGTGACCAATCCGCGGAGCGGCTGCGGCGCGGACGACGACCG	1741
Db	1468814	CCACCGTACCAATTGTGTCTTCAACCTGTCAAGGAGCGGCTGTGCGAGGAGATGAGACGG	1468755
Qy	1742	CGCGCTGTGTACGCGCGGCTTCCCGCGCGCTTCATGACCGGCGGCGCCCAAGAGCGCA	1801
Db	1468754	TGGACTGTGTATCGTGTCTGTCTCCCGGAGGTTTGTGTGACCGGAGCGCCGAAACTGTGCA	1468695
Qy	1802	CCATGAGATCATGACCGGCTTGAGGAGGCCCCCGGGGCGTCTACTCGGAGCGCTCG	1861

Db	1468694	CGATGGAGATCATCGACTCGCTGGAGACCGAGGCGCCCGGGGTGTACTCCGAGACCATCG	1468655
Qy	1862	GATGGTTGCGCCTCAGCGGCGCCCGACCTTCAGAGATGTATTCGACCATCGTGTGG	1921
Db	1468693	GCTACTTCGGGTGACGTGGGGCGCGGACCTCGCATTCGACGCGGACCGGCGTGTCA	1468575
Qy	1922	CCGACGCGCCAGGAGTTGCGCGTCGCGCGGGGAGATCGTGTCCCTCCGACCGAGGAGG	1981
Db	1468574	CCGACGGGGAGATGCACTTCGCGCGGGCGGGTGGATGTCTTGCTTCGATTCGGTTCG	1468515
Qy	1982	AGGAGTTCACGAGACCGTGTAAAGGCCCGCCATGTTCACCCCTCGACGGCAGCG	2041
Db	1468514	GCGAGTACGACGAGATGCTGCTGAAGACGCGCACCGAGATCGCGGCCACCGGAGACCGGA	1468455
Qy	2042	CCG 2044	
Db	1468454	TTCG 1468452	

```

RESULT 7
US-10-156-761-1
Sequence 1, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHITA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
Prior APPLICATION NUMBER: JP 2001-204089
Prior FILING DATE: 2001-05-30
Prior APPLICATION NUMBER: JP 2001-272697
Prior FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

	Query Match	14.6%	Score 301.4	DB:15	Length 9025608
	Best Local Similarity	49.7%	Pred. 10.4e-57		
	Matches	948	Conservative	0	Mismatches 936
				Indels	Gaps
				25	6
Qy	135	CGACCGCATGCTGTGTCCTCCCGGCGCCCGGAGCCCGACCGGGAACGGAGCTTCGGAAAT	194		
Db	7417356	CGGCACATAGTGTGGCATGATACACACACGCCCCAGCCGACGAGCGCCACAGACCTGTGACGC	7417415		
Qy	195	CAGCCGCGGGCGATACCGACGAGCGGCTCCGCTCTCGAGCGCTTCGCTCGGCGACCA	254		
Db	7417416	CGAAGATGGCGCGGGCCCTCAAGGCGAGCGCGAGGGGTCTCTCCCGCATGCGCCAGCA	7417475		
Qy	255	GGGCGATGGCCCAAGCTCTTCGGCGGAAACGTTGGCTTGCCCGGAAACCATGCAACGGCCG	314		
Db	7417476	GTACGACACCGGTGAGGGTCTGATGCTCGGTGATGATACGAGAGGCGATGACACGACAC	7417535		
Qy	315	GGTCTCCGAGGTGCGGCACACCGGCGAGAGCGTTCCTCGGGGCTTCGCCCTTGCGCTTAC	374		
Db	7417536	GCTACACACGGGCGCGCTGACCTACTGCTGCGGACGCGGGAGATGACTGGTGAAAGG	7417595		
Qy	375	CGCGGTGGCTTACACTCTCCCTGGCGGCGCACCGACCTCCCGGACGAGCTCGAACCCCTTCG	434		
Db	7417596	CGACACCTCGGGCCACGTCAGTACGTGAATCTGTAGCCCTTCGATCTGCAACGCCGACAC	7417655		

QY	435	CTGAGAGCAGCAGCGGGTTCGTCATCGGAGCTTCGGCACAACGAGAAAGCCGCTGTGGGGCGT	494
Db	7417656	CGTCTCGTCAAGGTTCACACAGGTGGGCGCCGCTGCACACACGGGCGCGACAGTGGCTT	7417715
QY	495	CCAGTTCACCCGGAGTCCATCGGCAGCGATTCGCGCGGAGATCATGCGCAATTCCTCG	554
Db	7417716	CGAGCGCAGACGTGTGCTCAAGAGCAGCGGATTCCTGGGTACCCGACGGGTATCAGTAAG	7417775
QY	555	CGACCTGCGCCCTGCCACCAACCGGGACCGGGCGCACCTCCCGTACGAACT	614
Db	7417776	TCA---GCGGCCATGACCTCGAGAGTTCGGAACCTGGCCACACGCGCGTGTAT	7417833
QY	615	CCACGTGCGCGCTCGACGTCGTCGCGCGCAGCGCAGGAAGGTATGCGCGCGCTCTGCC	674
Db	7417832	TCCGCTCAGCGCGGAAATCTCTCGCGCGACGGGAGAACCCCGTCCGGGCTTACCGGAACT	7417891
QY	675	CGGCGAGGGCACACGTTCTGAGCTGAGACAGAGCTCCGTCTCGAAGCGCGCTTCGCTT	734
Db	7417892	CGCGCGCAGAGCGCCCCCGGACACTTCTCTGAGATTCGCGGAGAACGCGCGCTCGGGCTT	7417951
QY	735	CTCCCTTCCTCGGACGACCGCGCGCGCGCTGCGCGAGTACTGACTTACACGCGGTGCGCA	794
Db	7417952	CAATGTGTCCTCGCTACTCTTCTGTGGGCGTTCGCTCCGCGCGCACTTACCGAGCGTGA	7418011
QY	795	CGGCGGTGTCTCGTCCGCGGCTTCGACGGGACACGACCCGAGCGGAGCGCGCCCTTCTT	854
Db	7418012	CGGGCAGGCGGACTGTGCTCGGACCCGCGCGTGGGCGTCCCCACGAGCGGACCGCGCT	7418071
QY	855	CAACTACTTGAAGAGACAGCTTCGAAAGCGCGACCGGATCCCCGTGCGCCCCCGAATGCGCTT	914
Db	7418072	CGCGCGCTCGCGCGCCACCATGAGGGCTGACACGCCCCACAGAGAGGCGCTTCGCGCC	7418131
QY	915	CGAGTTCAACTCGGCTACGTCGCGCTACTCTCGGCTACGA-----GCTGAAGCGGAGAC	968
Db	7418132	C---TTACCGGGCGGACATGTTCGGCTATCTTGGGTACGACATCGTGGCGCGCTTGAGAA	7418188
QY	969	CACCGGCGACCCCGCGCACCGGATCCCGCAACCCCGACGCGCGCTTCTTTCGCGCACCG	1028
Db	7418189	GATCGGCGCGCGGAGGCGACGATCTGAAACTTCCGAGCTGACATGCTCTTCAACAG	7418248
QY	1029	CGCATCGCCCTCGACACGACGAGGAAGGTGTGTACTCTGTGGCCCTTCGACCGCGGGG	1088
Db	7418249	CGATCTGCGCGTCAATGAGACACTTGGAGGGGCTCGTCTCTGTATCCCAACGCGATCAA	7418308
QY	1089	CCAAGACGAGCGGCGCCGCGCTGTGCTGCGGAGACGGCGGAGACCTTCAACCGGCTTGCG	1148
Db	7418309	CCACAAAGACCTCGACACGGGGCGTGCAGAGGCTTACGCGGAGCGCGTCCCGCTCGA	7418368
QY	1149	CGTCCGCGCGCCCGCGGCGGACCCCGCGCAtGTGTTTGGGATTCCTCGAGCGCGCGCG	1208
Db	7418369	CGCATGAGGCGGACCTGTTCGCGCG-----GTTGCGCAGCGCCCGCGCGCTTCCC	7418422
QY	1209	CGGCTTGGGCGCCCTTGCGCCGCGGCGGCAAGACGAGGAGCGCTACTCAAGCGCATCGA	1268
Db	7418423	GGCCTCGGAATCCTCGGAGTACACCGCGCTCTGTGGGGCGGCGCCGACTTCCAGACGCGCT	7418482
QY	1269	CGAGTGTCTCAAGAG---ATTCGCGCAACGGGAGTGTGTAAGATCTGCTGACCAAT	1325
Db	7418483	CGAGGACATCAAGAGAGCGATCCGGGCGGCGAGGCGTTTCAAGTGTCTTCCCGCGG	7418542
QY	1326	GGTCAACCGGCGGACGAGGCGGACCGCGCCGCGGCTTACTCGCGGTGTGGCGCGCATAG	1385
Db	7418543	CTTTCGAACGCGCGTGCACGGGCAAGCGCTTGAACGTCTTACGGGTACTTGGGGGACCA	7418602
QY	1386	CCCCGTCCCGTACGGGCGCCCTGTCTGAGTTCCCGAATGTGCGGTGTGACGCGCTCGCC	1445
Db	7418603	CCCCTCAACGTCATGTAATCTTTCGCTTCGAGGGCTTTCGAGTGTGCGGTCTGTCTCCC	7418662
QY	1446	CGAGCGGTTCTCAAGATGCGGCGGACACCGCGCGGCTGAGTCCAAAGCATCAAGGGAGAC	1505
Db	7418663	CGAGGCGCTCTGTCAAG---GTTCAGGACCGGCGAGGATGTACACCCATTCGCGGGCAC	7418719

QY	1506	CGCCCCCGGGGGCGGCACCGGGGGAGAGAGACAGCGCTCCGCGCCGACCTGGCCCCCGC	1565
Db	7418720	CCGGCCGGCGCGCGACCCCGCAGAGAGACAGGCCCTTCGCGACAGAGCTGTGGCCGA	7418779
QY	1566	GGAGAGAGACCGGGCCGAGAACTGATGATCGTCGACTGTGTCCGCAACGACTCAACAG	1625
Db	7418780	CCCCAAGAGCGCGCCGAGCACTGATGCTCGTCGACTGGGGGCGAACGATCTGGGGCG	7418839
QY	1626	CGTCGGCGGATTCGGCTCCGTCACAGTGCCTCCGCGCTTCGAGGTTGAGACTACGGGCC	1685
Db	7418840	GGTCGCGAGCCCGGTTCCGTCAGAGTGTGACTTATGACCAATCGAGCGCTACCTGCA	7418899
QY	1686	CGTCGACAGCTGTGTGACCATTCGGGGAGCGGCTGGGCGCCGCGACACGACGCGCGC	1745
Db	7418900	CGTATGACATCGTGTGACCGTGAACCGGGCGGGTCCGCGGGGGGGTACGCGCTTCGA	7418959
QY	1746	CTGCGTACGCGCGCGCTTCCCGCGCGGCTCCATGACCGGGCGCGCCCAAGAGCGACCAT	1805
Db	7418960	CGTGTGACGCGCTGTGCTTCCGGCGCGGACCTCTTCGGCGCGCCCAAGCGCCGCGCTT	7419019
QY	1806	GGAGATCATTCGACCGCTGTGAGAGAAAGCCCCCGGGGCGCTTAATCTCGGGGCGCTCGATG	1865
Db	7419020	GCAAGTCAATCGACGAACTGGAGCGGTCCCGGGGGGGGCGTGAACGGCGGCTGTGCGCTA	7419079
QY	1866	GTTGCGCCTTCAGCGGCGCGCGCGACCTGACATGTCTATCCGACCAATCGTGTGGCCGA	1925
Db	7419080	CCTTCGACTTTCGGGGCGACTCCGACACCGCCATCGCAATCCGACAGCGCGTGTGCGCCGA	7419139
QY	1926	CGGCGAGCGGAGTTCCGCGTCCGGGGGGCGCATCGTGTCCCTTCGACACGAGAGGAGGA	1985
Db	7419140	CGGCACGGCGTACGTGCAAGCGGGGCGCGGCACTGTCCGGAATCTCGAACCGGCTGCGCGA	7419199
QY	1986	GTTACCGGAGCCGTGTAAAGCGCCCGCCATGTGTCAGCGGCTTCGAC	2034
Db	7419200	GGACGACGAGTGGCGCAACGAGCGGCGCGCTCTGTGCGCGGCTTCAC	7419248

```

RESULT 8
US-10-156-761-6148
; Sequence 6148, Application US/10156761
; Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272657
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 6148
LENGTH: 1479
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1479)
US-10-156-761-6148

```

Query Match	13.6%	Score 280.8;	DB 15;	Length 1479;
Best Local Similarity	51.9%	Pred. NO. 5.1e-52;		
Matches 764; Conservative	0;	Mismatches 687;	Indels 21;	Gaps 5;

572 ACCACCGGGCACGGCGGCACGGGGCCGACTCCCGTAGCAACTCCACGTGCGCCGCTCG 631

Db	5	ACCTCGAAGCGTTCCGCAAGCTGGCCACCGACCGCGTGTCAATTCCGTGACCGGCAAC	64
Qy	632	ACGTGTCTCCGAGACGCCGAGAGAGTAAAGCCGCGGCTGTCTGCCCCGAGGCAACCACT	691
Db	65	TCTTCGCGGACGAGGACACCCCGGTGGGCTCTACCGCAAGCTGCGCGGAGCGCCCCG	124
Qy	692	TCTGGCTGGAAGAGCACTCCGTCTCTGAAAGGCGCTCGCGTTCTCTCTCTCGGCGACG	751
Db	125	GCACCTTCTCTTGGAGTCTCGGAGAAAGCGCGCTCGCGTTCAAAATGTCTCCGTACT	184
Qy	752	ACCGGCGCGCGCTCGCGAGTACCTTCACTCACCGCGTGCAGACGCGCGCTCTCCGTCC	811
Db	185	CTTTCGTGGGGGTCCGCTCCCGCGCCACCTTACAGAGCTGAGGGGAGGGCGCACTGCG	244
Qy	812	GCGGCTCGACGGGACACGACCCCGGACGCGCGCGCTTCTTCAACTACTGAGAGAGC	871
Db	245	TGGGACCCCGCGCGTGGGCGTCCCGACAGAGGCGACCGCGCTGCGCGCTGCGCGCA	304
Qy	872	AGCTTGAAAGCGGACAGGCTCCCCGTGCGCCCCGAACTGCTTGTGAAGTTCAACTCGGCT	931
Db	305	CCATCGAAGCGCTGACACAGCGCCACAGAGAGGCGCTGCGCGC---TTCACTCGGCGCA	361
Qy	932	ACGTGCGTACTCTGCGCTTACGA-----GCTGAAGGGGAGAACACCGGCGCACCCCGCGC	985
Db	362	TGTTGCGGTTACTCGGCTTACGACATCTGTGCGCGCGCTGGAAGAAATGGCCCCGCGAGC	421
Qy	986	ACCGGTCCCGCACCGCCGACGCGCGGTCTCTTTCGCGACCGGCGCACTGCGCTGAC	1045
Db	422	GCGACGATCTGAATCTCCGAGCGTACCAATGCTCTTACCAAGGAGATCTGGCGTCAATGG	481
Qy	1046	ACCGAAGAGGCTGCTGTACTCTGCTGGCGCTTGCACCGCGGAGGCAAGACGCGCGCC	1105
Db	482	ACCACTGGAGGCGTCTCGTCTGTGTATCGCAACGCGATCAACACACACACTTCACAA	541
Qy	1106	GCGCGCTGCTCGGAGAGCGGCGGACCCCTCAACGCGCTGGCGCTGCGCGCGCGCGCG	1165
Db	542	CGGCGTGTGACAGAGGCTTACGCGGAGCGCGGTGCGCTCGCTTGCAGCGCATGGAGCGGAC	601
Qy	1166	AGCGGACCCCGCATGTCTTTCGAGATCCCGGAGCGGCGCGCGCTTTCGCGCGCGCTGG	1225
Db	602	TGTGGCGGCG-----GGTGGGAGACCGCCCGCGCGCTTCCGCGCGCTCGAACTCCGCG	655
Qy	1226	CCCGCGCGCGCCACGACAGAGAGCGCTTACTCAAGCGATGACGAGTGTCTCAAGAG-	1284
Db	656	AGTACACCGCGCTCTGGGCGCGCCCGCACTTTCAGAGACGCGGTGAGGACATCAAGGAGC	715
Qy	1285	--ATTCGCAAGGCGGAGTGTACAGATCTGCGTGAACCAATGTGTACCGCGCGGACCG	1342
Db	716	GCATTCGGGCGGAGCGGCGTGTCCAGGTGTCTCCCTCCACGCGCTTGAAACGCGGTGCA	775
Qy	1343	AGGCGACGCGCGCTCGCTACTCTCCGCGCTGCGCGCATAGCGCGCTCCGCTACGGCG	1402
Db	776	CGGCAAGGCGCTTGGACGCTTACCGGGGTATCTGCGGGGACCAACCTCTCACCGTACATGT	835
Qy	1403	CCGTGCTGAAGTCCCGAATGTGTGCTGAGCGCTGCGCGGAGCGATTCTTCAAGA	1462
Db	836	ACCTCTTCCGCTTGAAGGCTTTCAGACGTGTGCGCTGCTCCCGAGGCGCTCTGTAAGG	895
Qy	1463	TGCGGCGCGAGCGGCGGTGCAAGTTCMAAGCCCATCMAAGGGAACCGCGCGCGGGGCGCA	1522
Db	896	TTCAG--GAGGGGACAGGTGTTCACCCCAATCGGGGACCCCGCGCGCGCGCGCA	952
Qy	1523	CGGCGAGAGAGAGAGCGGCGCTCGCGCGCGACCTGAGCGGCGGGAGAGAACCGGGCG	1582
Db	953	CCCGCGAGAGACCAAGCGCTTCCGCGGACGAGCTGTGCGGACCCCAAGAGCGCGCG	1012
Qy	1583	AGAACTTGATGATGTCGACCTGTGTGCGCAACGACTCAACAGCGTCTGCGCGATCGGCT	1642
Db	1013	AGACACTGATCTGTGACCTGGGGGGCAACGATCTGGGGCGGGTCTGCGAGCCGGTT	1072
Qy	1643	CCGTTCACGTGCCCCGCGCTTTCAGAGGTGAGACCTTACGCGCGCGGTGACCAAGCTGTGT	1702
Db	1073	CCGTTCAGAGTGTGTGACTTATGACATTCGAGCGCTTACTGCAAGTATGACATCTGTGT	1132

```

QY 1703 CGACCATCCGGGGAAGGCTGCGGCGCCGACACGACCGCGCTGCGTACGCGCGCT 1762
D 1133 CGACCGTGAACCGGCGGCGGCTGCGGCGGCGCTGACGCTTGAAGTGTGACGCGCTGCT 1192
QY 1763 TCCCGGCGGCTCCATGACCGGCGCGCCGAAAGGCGGACCGATGATGATGACCGCG 1822
D 1193 TCCCGGCGGCGGCTCCATGACCGGCGCGCCGAAAGGCGGCGGCTGATGATGATGACG 1252
QY 1823 TGAAGGAGGCGCGCGGCGCTGATGCTGCGGCGGCTGATGCTGCGCTGCGCTGCGCG 1882
D 1253 TGAAGCGGCGCGGCGGCTGATGCTGCGGCGGCTGATGCTGCGCTGCGCTGCGCG 1312
QY 1883 CCGCGGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1942
D 1313 ACTCGGACCGGCGGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATG 1372
QY 1943 GCGTGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2002
D 1373 AGCGGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1432
QY 2003 TAAAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2034
D 1433 ACAAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1464

```

RESULT 9

```

US-10-716-803-9
; Sequence 9, Application US/10716803
; Publication No. US20040229236A1
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; Crouzet, Noel
; Jacques, Nathalie
; Lacroix, Patricia
; Thibaut, Denis
; Zagorec, Monique
; Debussche, Laurent
; De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved in The
; Biosynthesis Of Streptogramins, Nucleotide Sequences
; Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegau, Henderson, Parabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/716,803
; FILING DATE: 20-Nov-2003
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 09/635,359
; FILING DATE: 09-AUG-2000
; APPLICATION NUMBER: US 09/231,818
; FILING DATE: 15-JAN-1999
; APPLICATION NUMBER: US 08/403,852
; FILING DATE: 10-MAY-1995
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Arisgo, Salvatore J.
; REGISTRATION NUMBER: 46,063

```

```

; REFERENCE/DOCKET NUMBER: 03806;0054-04000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.priestinaeapiralis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 61..645
; OTHER INFORMATION: /product= "gene papa"
; SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-10-716-803-9

```

```

Query Match 13.2%; Score 272.2; DB 20; Length 645;
Best Local Similarity 68.7%; Pred. No. 4.4e-50;
Matches 392; Conservative 0; Mismatches 173; Indels 6; Gaps 1;

```

```

QY 2 TGGCGACGCTTGTGATGACAACTACGACTGTTACCCAACTGTTCCAGTACATCG 61
D 71 TGGCGACGCTTGTGATGACAACTACGACTGTTACCCAACTGTTCCAGTACATCG 130
QY 62 GGGAGCGGACCGGCGGCAACCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 121
D 131 CGAGGTGAACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 190
QY 122 CGCTCGAG-----GACTTGAAGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 175
D 191 CCCTGGCGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 250
QY 176 GGGAGCGGACCTTGGGATGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 235
D 251 CGGACGCGGACCTTGGGATGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 310
QY 236 GCGTTCGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 295
D 311 GCGTTCGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 370
QY 296 CGGACCGGACCTTGGGATGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 355
D 371 CGGACCGGACCTTGGGATGACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 430
QY 356 GCGTTCGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 415
D 431 ACATTCGCTTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 490
QY 416 AGGAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 475
D 491 CGGAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 550
QY 476 AGGAGTGAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 535
D 551 ACCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 610
QY 536 AGATCATGCGCAACTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 566
D 611 GGATGCTGCGCAACTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 641

```

RESULT 10

```

US-10-282-122A-30135
; Sequence 30135, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangu

```

```

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlson, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trivick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EUTRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; PRIOR FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 30135
; LENGTH: 1362
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
; US-10-282-122A-30135

Query Match      13.1%; Score 270; DB 17; Length 1362;
Best Local Similarity 58.7%; Pred. No. 1.2e-49;
Matches 487; Conservative 0; Mismatches 340; Indels 3; Gaps 1;

QY      1227 CCGCGCGCGCCGACGACGAGAGCGCTTACTCAAGCGCGATGACGAGTGCCTCAAGAGAT 1286
DB      528 CCGACGAGCATGACGCGCATCCGACTATCGGACGAGCGATCCGCGCATCCAGAGACTACAT 587
QY      1287 CCGCAACGCGGATTCGACAGATTCGCTGACCAACATGATGACCGCGCGCGAGC 1346
DB      588 CCGAGGAGGAGCTGCTACCGAGTGAACATTAAGCCGCTTCCAGCCGCTTGACGCG 647
QY      1347 GACGCGCTTCGCTTACTCCGCGCTGCGCGCATGACCGCTCCGTCGCGCGCGCT 1406
DB      648 CTCGCGGTGCGCGCGCTATGCGCGCTGCGCGAGCGCTGCGCGCATCGCTTCTCCGCTA 707
QY      1407 GCTCGAGTTCCCGAACTGCTGCTGAGCGGCTTCGCGCGAGCGGTTCTCAAGATCG 1466
DB      708 CTGCGACTGCGCGAGCGCGCATCTCAAGCTGTGCGCGGAGCGCTTCTCAAGCTCG 767
QY      1467 CCGCGAGCGGCGGTGAGTCCAAAGCCATCAAGGGAGCCCGCGCGCGCGCGCGC 1526
DB      768 ---CAAGGGGCAAGTGGAAACCGCGCGCATCAAGGGAGCCCGCGCGCGCGCGCGC 824
QY      1527 GGAGGAGAGAGAGCGGCTCCGCGCGCGAGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAG 1586
DB      825 CGAGGAGAGAGATGCGCGCTGCGCGCGCTGCTGCGCGAGCGCGCGCGCGAGAGAGAG 884
QY      1587 CCGATGATGCTGAGACTGTGTCGCAAGCAAGACTCAAGCGCTGCGCGATCGGCTCCGT 1646
```

```

DB      885 CCGATGATGCTGAGCTGCTCGCGCAAGCAATCGGACCGAGTTGCGCAACTGCGAGCGT 944
QY      1647 CCACGTGCCCGCGCTTTGAGGTGAGAGCTAACGCGCGCGCGGACACCTGCTGCGAC 1706
DB      945 ACGGTAACGAGAGCTGTTTGCCCTGGAAAGCTTAATCCAAAGTGCATCACTGCTGAGCAG 1004
QY      1707 CATCCGGGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1766
DB      1005 CCGTAACGCGGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1064
QY      1767 CCGCGGCTTCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1826
DB      1065 CCGCGGCTGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1124
QY      1827 GGAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1886
DB      1125 ACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1184
QY      1887 CGACTCAGCATGCTGATCGGACCATCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1946
DB      1185 GAGACACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1244
QY      1947 CCGCGGCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2006
DB      1245 CCGCGGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1304
QY      2007 GCGCGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2056
DB      1305 GGTCCGGGTGCTGCTGAGAAACCTTGAGAAATGCGCGGAGACGCTCC 1354
```

RESULT 11

```

; US-10-501-282-17
; Sequence 17, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCWICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALLOIOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; FILE REFERENCE: AML00780 12
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 2046
; TYPE: DNA
; ORGANISM: Alloiococcus otitidis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (22) .. (2043)
; US-10-501-282-17
```

```

Query Match      13.0%; Score 267.2; DB 22; Length 2046;
Best Local Similarity 48.8%; Pred. No. 4.5e-49;
Matches 999; Conservative 0; Mismatches 963; Indels 84; Gaps 7;
```

```

QY      1 ATGCGAGCGCTCTGATGCAACACTAGCACTGTTACCCACAACTGTTCCAGTACATC 60
DB      22 ATGAACTCACTATTATGATATGATGATTTTATTACTTAACTTATACAGCTTATT 81
QY      61 GCGAGGCGCACCGGCGCAACCCCGCTGCTGTCGCCAAAGAGCGCGACTGTGCGGCGCTG 120
```


Db 82 GGGAGGTCAAGGAGAAAGCCCATGCTGATAAAAAATGACAGATGACTTACCAAGAA 141
Qy 121 CCCGTGCA---GGAATTGACGCGATGCTGTCGCCCGGGCCCCGGACAGCCCGA 177
Db 142 CTATTGATCTCGACTTTGATTAATGATATTATTTACACCGGGCCCGGATGTCGGACCG 201
Qy 178 GAACGGGACTTTGGAAATGACCGCGCGGCGATCAACGACGGGCTGCCGTCGCGC 237
Db 202 GACAAAGACTTCGGCTTGTCGGCGCAAGTAATGAAAAGCTTGACAAAGCTATCTTTGGC 261
Qy 238 GTCGCGCTCGGCGCAACGAGGCGATGCCAGCTCTTCCGCGGAACCGTCGCGCTGCCCG 297
Db 262 ATTTGCTCGGCGCAACGAGGCTTATTTATTAACCGTGCACCTTAATGCGGGCTGAT 321
Qy 298 GAACCCATGACGCGCGGGTCTCCGAGTGCAGCAACCGCGAGAGCTCTTCGCGGGC 357
Db 322 ATCCCATGCAATGCTCGGCAAGTTCGCTTACCATTAATGGAAAAATCTTTTGACGGC 381
Qy 358 CTCCTCGCGCTTACCGCGCTGCGCTACCTCCCTGG---CCGCGACGACCTCCCG 414
Db 382 ATGACCAAGGCTTTGAAGTCAACCGCTACCACTCTGTTGTGAAGACAAAGATTA 441
Qy 415 GACGAGCTCGAACCCCTCGCTGAGAGCAACGCGGCTGTCATGGCTCGGCGACGCG 474
Db 442 GAGAAATTTCAAAATGCAATGCCAAAACCGATGGAATGTGATGCGCTGCCCAAG 501
Qy 475 GAGAAAGCGGCTGAGGCGCTGCAAGTTCACCCGAGTCAATCGGACAGCACTTCGCGCG 534
Db 502 ACCAAGCCCATTAATGGGATCAATTCACCCGAAATCAATGCAACCGATATGGGAA 561
Qy 535 GAGATCAATGCACTTCGCGGACCTCGCGCTCGCGCAACGCGGAGCGGCGCAACGCG 594
Db 562 AAATGATGAAAACTT-----TATGCTTGTGCGCAGGAC 597
Qy 595 GCGGACTCCCCGTAACGACTCAAGTCCGCGCTGCAAGTGTGCTGCGGACCGCAAGAG 654
Db 598 TACTATTAACCAATCAAGCTTTTATTAATAAAAGTTCCGGAACTAGGAGCAACGCAAG 657
Qy 655 GTACGCGCGGCTGCGCTGCGCGGAGGAGCAACGCTTGGCTGAGCAGACGCTCCGTC 714
Db 658 CTCTACGATCAATTTGGCCAAATTTGATGACCAAGTACTTTGGCTGCACTCACAGAGTG 717
Qy 715 CTCGAAAGCGCTCGCGCTTCTCTTCTCGCGGACGACCGCGGCGCTCGCGAGTAC 774
Db 718 GACCGTGGCTTGTACGCTTTTTCATCTTTGCGATGCTGACCAAAAGCGGGCGCACACC 777
Qy 775 CTCACTACCGCGCTGCGGAGCGGCTGCTCCGTCGCGGCTCGGACGACACGACC 834
Db 778 TTAATAATACGATGTGCACAAAAAGAAAGTGAATAAGATAGCGTCCGACGAGTT 837
Qy 835 CGGACGCGGCGCTCTTCTTCAACTACCTGAGAGAGAGCTCGAACGCGAGCGGCTCCC 894
Db 838 GAAA---GCTTCCAGACTGACATCTTTCTTACTTAAAGCCAAACGACCAAAATGCGCT 894
Qy 895 GTCGCGCGCGAGCTGCGCTTGAAGTCACTCGGCTAGTGGCTACTCTCGGCTACGAG 954
Db 895 TATGAGCGCGCTTGGCTTGTACTTTCAGCTAGGCTATATGCTATATGCTGCTAGAG 954
Qy 955 CTGAAGCGCGAGAC---CACCGGCGACCCCGCGACCGGCTCGCGGACCCGACGCGCG 1011
Db 955 GTCAAAAAGATACAGTTCAAGTTCACCAACAGACAGTCACTTCGCGAGTGGCTAC 1014
Qy 1012 TTCTCTTGGCGGACCGCGCTGCTGACCAACGAGAGAGCTGCTGCTACCTGCTG 1071
Db 1015 TTACTATTATGTAACGCGGCTTGTGCTATGACACACGAGAGAGACTTCTTCTCTC 1074
Qy 1072 GCGCTGACCGCGGGGCGACAGACGCGGCGCGGCTGCGGAGAGAGCGCGAG 1131
Db 1075 TCTTACCA-----GATGAC 1089
Qy 1132 ACCCTCAACCGGCTGCGCGCTGCGCGCGCGCGGAGCGACCCCGCATGCTTCTCGG 1191
Db 1090 CAGGATTGATTCAAGGCTCAAAAGACAGGCTCAAGAGGCTCAAAAGTTACAGCAAGAA 1149

Qy 1192 ATCCCGAGCGCGCGCGCTTCCGCGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1251
Db 1150 GAGCAAGAGGCAAGCAAGACTTCCACGACTGCTTGG---TCAAGCAAGAGAGCGCC 1206
Qy 1252 TACCTCAAGGCAATGCAAGAGTCTCAAGAGATTCGCAACGCGGAGTCTGACGATC 1311
Db 1207 TATATCAAGACATTTGAAACCATCTCAAGACTGATCAAGGCTGGGAAAGTTATGAAGTT 1266
Qy 1312 TGCCTAACCAATGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1371
Db 1267 TGCCTAACCAACCGGCTGATATTTGAGGCTCAATTTGAGCTTCCCTTATTAACAACAC 1326
Qy 1372 CTGCGCGCATTCAGCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1431
Db 1327 TTCCGCGCAAGAAAGTCCCGCTGCAAGTCTGCTTCTTGGCTTGAATGATTTTCAAGTT 1386
Qy 1432 CTGAGCGCTTCCGCGCGGCTTCTCAAGATGCGCGCGCGCGCGCGCGCGCGCGCGCG 1491
Db 1387 CTTCCTCTTCCATGAGGCTTTTATTAATGTCGACAAAGAGACCGGCTTGTGACCACTAG 1446
Qy 1492 CCCATCAAGGAGACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1551
Db 1447 CCCATCAAGGAGGAGCGGCTCAGACGCGGAGACAGATGACCAAGAGACCAAGACTTGATGAA 1506
Qy 1552 GACCTGCGCGCGCGGAGAGACCGGCGCGAGAACCTGATATCTGTCACCTGCTCCG 1611
Db 1507 GGCCTCCGAGTGAAGAAACCAAGGCAAGAAACCTATATGATTTGCACTCTTGGCG 1566
Qy 1612 AACGACTCAACAGCTCTGCGGATCGGCTCGCTCAGCTGCGCGCGCGCTCTTCAAGT 1671
Db 1567 AATGATCTGGCGCTTTTGTGAAATCGGCTGCTGAGTACCAATGATGATGTA 1626
Qy 1672 GAGACTTACGCGCGCGCTGCAACAGCTGTGTGCAATCCGAGAGCGGCTGCGCGCGCG 1731
Db 1627 GAGACTTACCAACCTCCACAGCTGTGTACCAAGTTCAGGCGGCTCAAGAAAGAC 1686
Qy 1732 ACCAGACCGCGCGCTGCTGCGCGCGCGCTTCCCGCGCGCTGCTGACGCGCGCGCG 1791
Db 1687 CTAGATGTGTGAAGTGTGAAGAACACTTTCGCGGCGGCTGCACTGCGGCGCT 1746
Qy 1792 AAGAACGCAACCATGAGATTCATCGACCGCTGAGAGAGGCGCGCGCGCGCTCTACCT 1851
Db 1747 AAAAAAAGAACCTGGAATTTATGATGACTTGAAGAGCTTCCAGAGGATCTATTTCT 1806
Qy 1852 GGGGCGCTGAGTGTTCGCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1911
Db 1807 GGGACTATTGGCTTTTACCAACATTCACATATGACTTCAACATGCTCATCGGAGC 1866
Qy 1912 ATCTGCTGCGGAGCGCGCGGAGTTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCG 1971
Db 1867 AGTGTGTGAAGACAGACAAAGCAAGCATCGCGCTGCGGCGGCTGCTGCTATCT 1926
Qy 1972 GACCAAGAGAGAGTTCACCGAGACCGTGAAGGCGCGCGCGCGCGCGCGCGCGCG 2031
Db 1927 GATCTGAAGAGAGTGTGATGATGTTTAAAGCTTAAGGCGCGCTTGTCCGCTTG 1986
Qy 2032 GACGCG 2037
Db 1987 CAAGCC 1992

RESULT 12
US-10-501-282-6651/c
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCWIGHEAD, JOHN CALHOUN
; APPLICANT: ZAGORSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARLISH
; APPLICANT: FLETCHER, LEAH DYANE
; TITLE OF INVENTION: ALLOIDOCUS OTTIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF

FILE REFERENCE: AM100780 L2
CURRENT APPLICATION NUMBER: US/10/501,282
CURRENT FILING DATE: 2004-07-09
PRIOR APPLICATION NUMBER: 60/333,777
PRIOR FILING DATE: 2001-11-29
PRIOR APPLICATION NUMBER: 60/426,742
PRIOR FILING DATE: 2002-11-18
PRIOR APPLICATION NUMBER: PCT/US02/36123
PRIOR FILING DATE: 2002-11-25
NUMBER OF SEQ ID NOS: 6653
SOFTWARE: PatentIn version 3.2
SEQ ID NO 6651
LENGTH: 1754382
TYPE: DNA
ORGANISM: *Alloioococcus otitidis*
US-10-501-282-6651

Query Match 13.0%; Score 267.2; DB 22; Length 1754382;
Best Local Similarity 48.8%; Pired. No. 1.6e-49;
Matches 999; Conservative 0; Mismatches 963; Indels 84; Gaps 7;

QY 1 ATGCGACGCTTCTGATCGACAACTACGACTCGTTACCCCAACCTGTTCCAGTACATC 60
DB 7807 ATGAAGTACTCATATGATGATACTATGATTCTTATACCTTAACTTATACAGCTTATT 7748
QY 61 GCGAGGCGCAACCGGCAACCCCGTGTGCGCCCAAGCGCGCACTGTGCGGCTG 120
DB 7747 GCGAAGGTACAGGGAAGAGCCCATGTGTATAAAAATGACAGATGACTACAAAGAA 7688
QY 121 CCCGTGCA--GGACTTGAAGCGAGTCGTGTCCCGGAGCCCGGACGCCGCAACGG 177
DB 7687 CTATTGATCTCGACTTTGATTAATGTCATTATTTTCAACCGGGGCCCGGTAGTCGGAACGG 7628
QY 178 GAACGGACCTTCGGAATCACGCCCGCGGCGATCACCGACAGCGCCCTGCGCTCTCGGC 237
DB 7627 GACAAGACTTCGCGTGTGTCCGCGAAGTAATGAAAAGCTTGAACAAGCTTATCTTTGGC 7568
QY 238 GTCTGCTCGGCGCAACGAGGCATCGCCAGCTCTTCCGCGAAACCGTCCGCTCGCCCG 297
DB 7567 ATTTCCTGCGGCGCAACGAGGCACTTATTTATTAACGCTGCGCACTTAAAGTGGGGCTGAT 7508
QY 298 GAACCCATGACAGCGCGGCTCTCCGAGTGCAGCAACCGCGAAGAGAGCTTTCGCGGCG 357
DB 7507 ATCCCCATGATGATGCGGCGAGAGTTCGCTTACCATTAATGGAAAAATATCTTTGACGGC 7448
QY 358 CTCCCTCGCGCTTACCGCGGTGCGCTTACCTCTCCCTG--CCGCGACCGACTCCCG 414
DB 7447 ATAGACCAAGCTTTGAAGTCAACCGCTACCACTCATTTGTTGTGAAGCAAGAAATTA 7388
QY 415 GACGAGCTCGAAACCCCTCGGCTGAGGAGCAAGCGGGGTGCTCATGGGCTGCGGCAACCGC 474
DB 7387 GAGAAATTTCAATTCATGTCACCAAAACCGATGATGAAATTTGTCAATGGCCCTGTCCCAAG 7328
QY 475 GAGAAAGCGGTGTGGGGGCTTCAGTTTCACCCGAGTTCATCGGCAAGCACTTCGCGCGG 534
DB 7327 ACCAAGCCCATATGAGGTTCATATTCACCCGAAATTCATGTCACCGAATGAGGAA 7268
QY 535 GAGATCATGGCCACTTCCGCGACCTGCGCTGCGCCCAACCGGGGAGCGGCGCCACGGG 594
DB 7267 AAATATGTGAAAACTT-----TATGGCTTTGTGCAAGAGAC 7232
QY 595 GCGGACTCCCGTATGCAACTCAAGTGCAGCGGCTGAGCGTGCAGCGCGCGCAAGAG 654
DB 7231 TACTATTAACCAATTCAGCTTTATTTATGAAAAAGTTCCGGAACTAGGGAACAGCAAGAC 7172
QY 655 GTACGCGCGAGCTGCTGCGCGGAGGAGCAACGTTCTGTGCTGAGCAGCAGCTCCGTC 714
DB 7171 CTATAGATCAATTTGGCCAAATGATGACCAAGTACTTTGGCTGCTCCAGAGAGTG 7112
QY 715 CTGGAAGCGGCTTCTCTCTCTCTGCGGAGCAACCGGCGCGCTGCGCAATAC 774
DB 7111 GAGCTGCTTGTTCACGAGTTTTCATCTTTGCGATGCGTGGACCAAGCGGGGCGACACC 7052

QY 775 CTCACCTACCGGCTGCGCGAGCGCTGCTTCCTCGTCCGCGGCTCCGACGGCACACGACC 834
DB 7051 TTAAAAATAGATGTGACCCAAAAAGAGTGAAGAAAAAACAAGATAGACAGTCCGACAGATT 6992
QY 835 CGAGCGGCGGCGCTTCTTCAACTACCTGAGAGAGCAGCTGGAAGCGCGAGGCTCCCG 894
DB 6991 GAAA--GCTTCCAAAGCTGACATCTTTTCTTAAAGGCCAACCGCAAAATGCGCT 6935
QY 895 GTGCGCCCGCACTGCGCTTCGAGTTCAACCTCGGCTACGTGCTACCTCGGCTACGAG 954
DB 6934 TATGAGCGGCGCTTGGCTTTTGAATCTTTTGAAGCTTATGAGCTTATGCGCTATGAG 6875
QY 955 CTGAAGCGGAGAC--CACCGGCAACCCCGGCAACCGGTCCCGCACCCGACCGCGCG 1011
DB 6874 GTCAAAAAAGATACAGTTCAAGTCCACAAACAGACACAGTCAACCTATCCGATGCTAC 6815
QY 1012 TTCTCTTGGCGGACCGGCGCATGCGCTTCGACACACAGAAAGCTGTGCTACTCTGCTG 1071
DB 6814 TTTACTTATTTGACCGGCGCTTGGTCTATGACACACAGAAAGAGACCTTTACTTCTCG 6755
QY 1072 GCGCTGACCGCGGGGCGACGACGCGCGCGCTGCGTGGGAGACGGCGGAG 1131
DB 6754 TCTTACCA-----GGATGAC 6740
QY 1132 ACCCTCACCGGCTGCGCTGCGCGCCCGCGAGCCGACCCCGCATGATCTTTCGGG 1191
DB 6739 CAGGATGGAATTCAGCGGCTCAAAAGCAGGCTCAAGCAGGCGAGTCAAGTTGACAAAGA 6680
QY 1192 ATCCCGAGAGCGGCGGCTTTCGAGCCCTTCGCGCGCGCGCGCAAGCAAGAGCGC 1251
DB 6679 GGAACAAAGGAGCAAGCAACTTCCACGACGCGGCTTCG--TCAAAGACAAACAGCGC 6623
QY 1252 TACCTCAAGCGGATGACAGAGTGCCTCAAGAGATCCGCAAGCGGCGAGTCTGACAGATTC 1311
DB 6622 TATATCAAGACATTTGAAACATTCACAGCTGATCAAGGCTGCGGAAAGTTATGAAGTT 6563
QY 1312 TGCCTGACCAATGTGTCAACGCGCGACCGAGCGGCGCGCTGCGCTTATCTCCGCG 1371
DB 6562 TGCCTGACCAACCGGCTGATATTGAGGGTCAAGATTGATGCGCTGCTTATTAACAACAC 6503
QY 1372 CTGCGGCGATCAGCCCGCTTCCGTAAGGCGGCTGCTGAGTTCCCGAATCTGCGG 1431
DB 6502 TTGCGCGAAGAAAGTCCCGCTGATGATCTGCTTTTTCCTTTGATGATATTTCAAGTT 6443
QY 1432 CTGAGCGCTGCGCGAGCGGTTCCCTACAGTCCGCGCGGCGGCGGCTGAGTCAAG 1491
DB 6442 CTTTCTCTTCAATGAGCGTTTATTAATGTCAGACAGGACCGGCTTGTGACACTAG 6383
QY 1492 CCATCAAGGAGACCCGCGCGGCGGCGCAACCGGAGAGAGAGAGCGGCTCCGCGCG 1551
DB 6382 CCATCAAGGAGAGCGGTCAAGAGGCGGGAACAGATGACAAAGAACCAAGACTTGAATAGA 6323
QY 1552 GACCTGCGCGCGGAGAGAGACCGGCGCGAAGACTGATGATGTGCACTGTGCTCGC 1611
DB 6322 GGCCTCGGAGTGAAGAGAAACCAAGGCGAAGAACTGATATTTGTCACCTCTTGC 6263
QY 1612 AACGACTCAAGAGGTCTGCGCGATTCGCTCGTCAAGTCCCGGCTCTTCAAGG 1671
DB 6262 AATGATCTGAGCGGTTTGTGAATCGGCTGAGTGAAGTACCAAACTGATGATGATGA 6203
QY 1672 GAGACTTACGCGCGCGTGCAGCAGCTGATGACCACTCCGAGAGCGCTGCGCGCGCG 1731
DB 6202 GAGACTTACGCGCGCTGCAGCAGCTGATGATGATGATGATGATGATGATGATGATGAT 6143
QY 1732 ACCAGACCGCGCTGCGTACGCGCGCTTCCCGGCTCCATGACCGGCGCGCG 1791
DB 6142 CTAGATGTGATGAGTGTGAAGAAACACTTCCCGGCGGCTGATGATGATGATGATGAT 6083
QY 1792 AAGAAAGCAAGATGAGATTCATGACCGGCTGAGAGAAAGGCGCGGCGGCTTACTTC 1851
DB 6082 AAAAAAGAACCTGGAATTTATGATGATGATGATGATGATGATGATGATGATGATGAT 6023
QY 1852 GGGGCGCTCGAGTGTTCGCTTCAAGCGCGCGCGCGCACTGAGATGTCATCGCAC 1911

Db 6022 GGGACTATTTGCTTTTAAAGCCAAATTCACATTAAGCTTCAACATGTCATCCGAGCC 5963
Qy 1912 ATCTGTGTCGGCCGAGCGCCAGAGTTCGCGCTCGCGGGGAGTCGTCTCCCTCC 1971
Db 5962 AGTGTGTGGAAGCAGACAGGCAAGCATCGGCTCGGGGGTGTGATGTCATGCTATCT 5903
Qy 1972 GACCAGAGAGAGATTCAAGCAGACCGTGTAAAGCCCGCCGATGTCACCGCCCTC 2031
Db 5902 GATCCTGAAGAGAGTTTGAATGAGTTTAAAGCTAAGGGGCTTGTCCGCTTG 5843
Qy 2032 GACGCG 2037
Db 5842 CAAGCC 5837

RESULT 13
US-09-738-626-1111
; Sequence 1111, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1111
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: *Corynebacterium glutamicum*
US-09-738-626-1111

Query Match 12.9%; Score 266.8; DB 9; Length 1860;
Best Local Similarity 59.7%; Pred. No. 5.6e-49;
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
Qy 1242 CAAGAGCGCTTACTCAAGCGCATCGACGATGCTCAAGAGATCCGACACGGCGAGTC 1301
Db 1059 CAAGATGAGTATCTCGAATAATTGCGAGGCCGAGAGCTGATTACTCGCGGCAATC 1118
Qy 1302 GACGAGATCTGCTTACCAACATGCTACCGCGCGGACGAGCGACGGCTGCGCT 1361
Db 1119 GTATGAAATCTGCTACCAACAACCTTCAAGGCAACACTGATGTCCTCTGCTGCTGC 1178
Qy 1362 CTACTCGCGCTCGCGCCATCGAGCCGTCCTGTCAGGCGCGCTGTCGATTCCTCCCA 1421
Db 1179 CTATCTAGACACTGCGTGGGCGCATCCACCGCATATGCTATCTTCAGTGGGGGA 1238
Qy 1422 ACTGTGGTGTGAGGCGCTCGCGCGAGGCTTCTCAGATCGCGCGGACGGCGCGCT 1481
Db 1239 TACCTCTATTTGAGTTCTCGCGGAGGTTTCATCATTGATTCGGCAGGATATCT 1298
Qy 1482 CGAGTCAAGCCCATCAAGGGACCGCCCGGGCGGACCGCGGAGAGAGCAGAGCG 1541
Db 1299 GGATTCAAAGCCCATTAAGGACACGAGCGGCTGGGCGAAGCGCAAGAGACCAAGA 1358

Qy 1542 GCTCCGCGCGGACCTGCGCGGCGGAGAGACCGGCGGAGACCTGATGATCGCA 1601
Db 1359 AATCATTTGCTGACTGCGAGTAACTCTTAAAGATCTGTGAGAAACTTATGATCTGGA 1418
Qy 1602 CTGTGTCCGACAGACTCAACAGCGTCTGCGGATCGGCTCCGTCAACGTCCCGGCT 1661
Db 1419 TTTGTGTCCGACAGACTTAAAGCCCGCGCGCTTTGCCACACAGTTAAACATCAAGCT 1478
Qy 1662 CTTCGAGTGAAGACTTACGCGCGCTGACACAGCTGTGTGACCATCCGGAGCGCT 1721
Db 1479 TTTGACGTGGAACCTTACGCGCACGTCCACCACTTGTACGACCGCTCTGCAAGATT 1538
Qy 1722 GCGCGCGGACGACGACCGCGCTCGGTCGAGGCGCGCTTCCGCGGCGCTCAATGAC 1781
Db 1539 GGGGCCACG---CAGTCCGATTGAGTGTGCGCGAGCATTCGCCGATGTCATGAC 1595
Qy 1782 CGGCGGCGCCCAAGAGACGACCATGAGATCATGACCGCTGAGAGAAAGCCCGGAG 1841
Db 1596 TGGTCCCGCAAGCTGCGACCATGAGATCATGATGAGCTGAGAGGCTCTGCGCG 1655
Qy 1842 CGTCTACTCGGGGCGCTGATGCTTCCGCTCAAGCGCGCGCGACCTCAGATCGT 1901
Db 1656 TATTTACTGAGGTGCTGAGATATTTTCCCTCGACGCGCAGTTGATCTCTCATGGT 1715
Qy 1902 CATCCGACCATGCTGCTGCGCGAGCGGAGGCGGATTCGCGGCGCGCGCATCGT 1961
Db 1716 GATCAGAACTCTGATATCAGAAACATCATCTGAGTACGAGTGGGCGGTCACTTCT 1775
Qy 1962 GTCCCTCTCCGACAGAGAGAGATTCAACGAGACCTGTAAAGGCCCGCGCATGCT 2021
Db 1776 TGCTGTGTGATCCGAGAGCTGAGTGGAGAAATCCCGTTAATACGCGCTCTGCT 1835
Qy 2022 CA 2023
Db 1836 GA 1837

RESULT 14
US-10-494-675-25
; Sequence 25, Application US/10494675
; Publication No. US20050019877A1
; GENERAL INFORMATION:
; APPLICANT: Zelder, Oskar
; APPLICANT: Pompejus, Markus
; APPLICANT: Schroder, Hartwig
; APPLICANT: Krogger, Burkhard
; APPLICANT: Klopprogge, Corinna
; APPLICANT: Haberhauer, Gregor
; TITLE OF INVENTION: Genes coding for metabolic pathway proteins
; FILE REFERENCE: BGI-163US
; CURRENT APPLICATION NUMBER: US/10/494, 675
; CURRENT FILING DATE: 2004-05-04
; PRIOR APPLICATION NUMBER: PCT/EP02/12141
; PRIOR FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: DE 101 54 292.1
; PRIOR FILING DATE: 2001-11-05
; NUMBER OF SEQ ID NOS: 164
; SEQ ID NO 25
; LENGTH: 2005
; TYPE: DNA
; ORGANISM: *Corynebacterium glutamicum*
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)..(1975)
; OTHER INFORMATION: RXA00579
US-10-494-675-25

Query Match 12.9%; Score 266.8; DB 21; Length 2005;
Best Local Similarity 59.7%; Pred. No. 5.5e-49;
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;
Qy 1242 CAAGAGCGCTTACTCAAGCGCATCGACGATGCTCAAGAGATCCGACACGGCGAGTC 1301

Db 1174 CAAAGATGAGTATCTGCACAAAATTCGACAGGCCAGAGACTGATTACTCGCGCGAATC 1233
QY 1302 GTACGAGATCTGCTGCTGACCAACATGGTCACTCGGCGGACCGAGCGGCGCTGCGCGCT 1361
Db 1234 GTATGAATCTGCTGCTGACCAAACTTCAGGGGACACACTGATGTGCCCCCTCTGGTGC 1293
QY 1362 CTACTCCGCGCTGCGGCGCATCAGCCCGTCCGTAACGCGCCCTGCTCGAGTTCCCGA 1421
Db 1294 CTATCTAGCACTGCGGGGCGCAATCCACCGCATATGGTGGCTATCTTCACTGAGTGGGG 1353
QY 1422 ACTGTGGTCTGAGCGCTCTGCGCCGAGCGGTTCTTCAGCATGGCGCGCGAGCGCT 1481
Db 1354 TACCTTATTTTAAAGTTCCCTGCGGAGCGGTTCACTACATTTGATTCGCGAGGTAATG 1413
QY 1482 CGAGTCCAAAGCCCATCAAGAGGAGACCGCGCCCGGGGGGCGACCGCGGAGGAGACGAGCG 1541
Db 1414 GGATCAAAAGCCCATTAAGAGCCACAGGCGCGCTGGCGCAACGCGGAGAGAGACCAAGA 1473
QY 1542 GCTCCGCGCGCATCTGCGCGCGGCGGAGAGACCGGCGCGAGAACCTGTATGATCGTCA 1601
Db 1474 AATCATTTGCTGAGCTGCGGAGTAATCTTAAAGATCGTGAGAGAAAATTGATGATCGTGA 1533
QY 1602 CCGTGTCCGCAAGACCTCAACAGCGTCTGCGGATGCGCTCCGTCCAAGTCCCGCGCT 1661
Db 1534 TTTGTGTCGCAAGCACTTACCTCGCGCGCTTTTGCCACCAAGTTAAACATTCAGAGCT 1593
QY 1662 CTTGAGGTGAGACCTGACGCGCGCGGCGGAGACGAGTGTGTGACCATCCGGGAGCGGCT 1721
Db 1594 TTTGACGTGAGAACCTTACGCAAGTCAACAACTTGTACGACCGTCTCTGCAAGATT 1653
QY 1722 GCGGCGCGGACCAAGACCGCGCTGCTGACGCGCGCGCTTCCCGCGGCTTCATGAC 1781
Db 1654 GGGGCCACG---CAGTCCGATTGAGTGGCGGCGGAGCATTCGCGGTGTTGATGAC 1710
QY 1782 CCGCGCGCCCAAGAGCGCACCTGAGATCATGACCGCTTGAAGAGGCCCGCGG 1841
Db 1711 TGGTGGCCCAAGCTGCGGACATGAGATCATGACTGAGAGCGACTCTCTGCGG 1770
QY 1842 CGTCTACTCGGGGCGCTGAGTGTGCTGCGCTCAGGCGGCGCGGACCTCGACATCGT 1901
Db 1771 TATTTACTCAGGTGGCTTGGGATATTTTCTCTGACGGGCGGAGTTGATCTTCATGCT 1830
QY 1902 CATCCGACCATCTGCTGCGCGACGCGGCGGAGTTGCGGCTGCGCGGCGCATGCT 1961
Db 1831 GATCAAGACTCTGCTGATCCAGAAATACGTTGAGTACGAGTGGGCGGTGCTACTTCT 1890
QY 1962 GTCCCTCTCGGACGAGAGAGAGTTCACCGAGCCGTGTAAAGGCCCGCGCATGCT 2021
Db 1891 TGGTGTGTGATCCGAGGCTGAGTGGAGGAAATCCGCTTAAATCAAGGCTCTGCT 1950
QY 2022 CA 2023
Db 1951 GA 1952

RESULT 15
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626

; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Query Match 12.9%; Score 266.8; DB 9; Length 3309400;
Best Local Similarity 59.7%; Pred. No. 1.8e-49;
Matches 467; Conservative 0; Mismatches 312; Indels 3; Gaps 1;

QY 1242 CAAAGAGCGCTACTCTCAAGCGCATCGAGAGTGCCTCAAGAGATCCGAAACGCGAGTTC 1301
Db 1053079 CAAAGATGAGTATCTGACAAAATTCGACAGAGCCAGAGCTGATTAATCTCGGCGAATC 1053138
QY 1302 GTACGAGATCTGCTGACCAACATGGTCAACCGCGGACCGAGCGGCGCGCTGCGCT 1361
Db 1053139 GTATGAATCTGCTGCTGACCAAAAATTCAAGGACACACTGATGTGGCCCTCTGGCTGC 1053198
QY 1362 CTACTCCGCGCTGCGCGCATGACGCCCTCCGTAACGCGCGCTGCTGATTCGCCGA 1421
Db 1053199 CTATCTGACCTGCGGGGCGCAATCCACCGCATATGTGTGTATCTTCACTGGGGGA 1053258
QY 1422 ACTGTGGTCTGAGAGCGCTCGCGCGGAGGTTCCCTCAGATCGGCGCGGCGCGCT 1481
Db 1053259 TACTCTATTTTGAAGTCTCTCGCGGAGCGTTTATACATTTGATTCGCGAGGTAAT 1053318
QY 1482 CGAGTCCAAAGCCCATCAAGAGGAGACCGCGCCCGGGGCGGACCGCGGAGAGAGACGAGCG 1541
Db 1053319 GGATCAAAAGCCCATTAAGAGCACGAGCGCGGTGGGCGAACGCGCAAGAGACCAAGA 1053378
QY 1542 GCTCCGCGCGGACCTGAGCGCGCGCGGAGAGAGACCGGCGCGGAAACTGATGATGCTCA 1601
Db 1053379 AATCATTTGCTGAGCTGCGGAGTAATCTTAAAGATCGTGAAGAAAATTGATATGCTGGA 1053438
QY 1602 CCGTGTCCGCAAGACCTCAAGAGCGTTCGCGCATGCGCTCGTCCAGCTGCGCGGCT 1661
Db 1053439 TTTGTCCGCAAGACTTAGCCCGGCGCTTGGCCACCAAGTTAAACATTCAGAGT 1053498
QY 1662 CTTGAGGTGAGACTTACGCGCGCGTGCACCAAGCTGTGTGATTCAGGAGACGCT 1721
Db 1053499 TTTGACGTGAGAACTTACGCGCACAGTCCACCAATTGTCAACACGCTCTGAGAGATT 1053558
QY 1722 GCGGCGCGGACCAAGACCGCGCTGCTGAGCGCGCGCTTCCCGGCGGCTCTCATGAC 1781
Db 1053559 GGGGCCACG---CAGTCCGATTGAGTGGCGGAGGATTCGCCGAGTTCGATGAC 1053615
QY 1782 CCGCGCGCCCAAGAGCGGACCATGAGATCATCAGCGCTGAGAGAGGCCCGCGG 1841
Db 1053616 TGGTCCCAAGAGCTGCGACATGAGATCATGATGAGCTGAGAGGACGCTCTGCGG 1053675
QY 1842 CGTCTACTCGGGGCGCTGAGTGTGCTGCGCTCAGCGGCGCGCGCATGAGATGCT 1901
Db 1053676 TATTTACTCAGGTGGCTTGGGATATTTTCTCTGACCGGCGGAGTTGATCTTCATGCT 1053735
QY 1902 CATCCGACCATGCTGCTGCGGAGCGGCGGAGGTTGCGGCTGCGGCGGCGATGCT 1961
Db 1053736 GATCAAGACTCTGCTGATCCAGAAATACCTGAGTACGAGTGGGCGGTGCACTTCT 1053795
QY 1962 GTCCCTCTCGGACGAGAGAGAGTTCACCGAGCCGTGTAAAGGCCCGCGCATGCT 2021
Db 1053796 TGGTGTGTGATCCGAGGCTGAGTGGAGGAAATCCGCTTAAATCAAGGCTCTGCT 1053855
QY 2022 CA 2023

Thu Oct 6 09:19:05 2005

us-10-089-514-1.rnpb

Page 17

Db 1053856 GA 1053857

Search completed: October 5, 2005, 09:56:39
Job time : 2072.38 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:20:08 ; Search time 7793.81 Seconds
(without alignments)
10665.737 Million cell updates/sec

Title: US-10-089-514-1

Perfect score: 2061
Sequence: 1 atgcgcacgcttcgtacga.....ccgtggcggcgcccgatga 2061

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: gb_est1.*
2: gb_est2.*
3: gb_est3.*
4: gb_est4.*
5: gb_est5.*
6: gb_est6.*
7: gb_est7.*
8: gb_est8.*
9: gb_est9.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216.6	10.5	529	CA003778	CA003778 HS15114r
2	212.8	10.3	845	BZ570522	BZ570522 msh2_1411
3	205.8	10.0	653	CF863831	CF863831 p82S008xg
4	195.6	9.5	1043	BZ548859	BZ548859 pacc1-60
5	178	8.6	523	BZ412600	BZ412600 OGAAG11TM
6	178	8.6	631	BZ412593	BZ412593 OGAAG11TC
7	178	8.6	631	BZ537360	BZ537360 OGAAG11TC
8	178	8.6	810	CG170785	CG170785 PUIR587D
9	174.2	8.5	579	CF598037	CF598037 NCEST3e06
10	171.8	8.3	660	BZ892725	BZ892725 Hm7_0152.
11	167.8	8.1	894	CC024641	CC024641 EST803025
12	167.2	8.1	757	BZ896109	BZ896109 NARP6_013
13	164.8	8.0	750	CC028609	CC028609 EST806993
14	164.8	8.0	818	CC030342	CC030342 EST808726
15	164.8	8.0	820	CC034510	CC034510 EST812894
16	163	7.9	768	CC0980064	CC0980064 GMB9007A1
17	162.8	7.9	825	CC144065	CC144065 NDL_75A11
18	162.8	7.9	859	BZ658136	BZ658136 GMB700005A
19	160.6	7.8	1674	CL078342	CL078342 CH216-149
20	159.6	7.7	2041	AG365808	AG365808 Mus muscu
21	158.6	7.7	1259	BZ557654	BZ557654 pacc1-60
22	158.4	7.7	432	CV122440	CV122440 Md1v4003F
23	151.2	7.3	1695	CC290874	CC290874 CH261-172
24	150.6	7.3	1031	BZ561643	BZ561643 pacc2-164

25	150.4	7.3	492	6	CB518127	CB518127 EST01 DNA
26	148.2	7.2	528	2	AW509018	AW509018 s139b01.Y
27	148.2	7.2	698	5	B0625203	B0625203 USDA-PP 0
28	147	7.1	1628	9	CG757066	CG757066 P052-2-1A0
29	145.4	7.1	1450	9	AG429738	AG429738 Mus muscu
30	143.8	7.0	1232	9	AG363333	AG363333 Mus muscu
31	141	6.8	1821	9	CL090560	CL090560 ISB1-17N1
32	139.6	6.8	1523	9	AG448267	AG448267 Mus muscu
33	139.2	6.8	1567	6	CG746709	CG746709 P040-1-B0
34	139	6.7	716	6	CD407015	CD407015 Gm_CK3199
35	138.8	6.7	1956	9	CG754548	CG754548 P050-1-A1
36	138.4	6.7	565	6	CA640597	CA640597 wren1n.pk0
37	138.4	6.7	669	4	BI955833	BI955833 HVMEM002
38	138.2	6.7	1822	9	AG435170	AG435170 Mus muscu
39	138.2	6.7	2243	9	AG381986	AG381986 Mus muscu
40	137.6	6.7	526	7	CF887023	CF887023 UI-CF-D01
41	136.2	6.6	540	4	BI427109	BI427109 sah76601.
42	136.2	6.6	1793	9	CG754612	CG754612 P050-1-D0
43	136	6.6	1738	9	CG750956	CG750956 P045-2-H0
44	136	6.6	1830	9	CL078620	CL078620 CH216-151
45	135	6.6	1642	9	CG754720	CG754720 P050-2-A0

ALIGNMENTS

RESULT 1
CA003778
LOCUS HS15114r 529 bp mRNA linear EST 23-OCT-2002
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION CA003778
VERSION CA003778.1 GI:24280760
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae; Hordeum.
1 (bases 1 to 529)
Zhang,H., Potokina,E., Michalek,W., Weschke,W., Stein,N. and Graner,A.
REFERENCE
AUTHORS Barley ESTs from germinating seeds
TITLE Unpublished (2002)
JOURNAL Contact: Stein Nils
COMMENT Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert length: 529 Std Error: 0.00
Plate: 15 row: 1 column: 14
Seq primer: M13rev
Location/Qualifiers
1. 529
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:258990"
/db_xref="taxon:112509"
/clone="HS15114"
/issue_type="embryo + scutellum"
/dev stage="0-16 hours after imbibition"
/lab_host="X110-Gold"
/clone_lib="HS"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also

due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN

Query Match 10.5%; Score 216.6; DB 6; Length 529;

Best Local Similarity 63.1%; Pred. No. 3,7e-27;

Matches 333; Conservative 0; Mismatches 195; Indels 0; Gaps 0;

```
QY 1482 CGAGTCCAGCCCATCAAGGGAGACCCGCCCCGGGGGGGACACCGGAGAGAGACGAGC 1541
    |||
DB 1 CGGACAGAGGCCCAATCAAGATACATAGACAGCTGGCAGAAACACAGAGAGAAATGAGT 60
    |||
QY 1542 GCTCCGCGCCGACCTGGCCGCGGAGAGAGACCGGGCCGAGAACTGATGATCTGCA 1601
    |||
DB 61 TCTACGTTTCACTGAATATACAGTGAAGAGACAGAGCTGAGAACTTGATGATTTGTA 120
    |||
QY 1602 CTTGTCGCGACAGACTCTCAACAGCGTTCGCGGATGGCTCCGTCCAGCTGCCCGGCT 1661
    |||
DB 121 TCTCTTAAGAAATGATCTGGGTAGGCTCGAGCCTGGAGCGGTGATGTTCTCGCCT 180
    |||
QY 1662 CTTGAGAGTGAAGACCTAGCGCGCCGCGACAGCTGTGTGACCATCCGGGGAGCGCT 1721
    |||
DB 181 CATGAGAGCTGGAATCATATATATCTGTTCAACCATGATGATGACCATCCGCGAACA 240
    |||
QY 1722 GCGGCGCGGACAGACACGCGCGCTGCGTACGCGCGCCTTCCGCGCGCTCATGAC 1781
    |||
DB 241 GAAGCTGATCTAAGCCCGGTAGACTGATCAAGAGCCGCTTCCAGAGAGGCTCATGAC 300
    |||
QY 1782 CCGCGCGGCCCAAGAGGACCATGAGATCATGACCGCCTTGAGAGAGGCCCGCGGG 1841
    |||
DB 301 GGGGTGCCCCCAAGGTCAAGTCAATGAGATCTCGACGCTTGAAGAGCAAGCCCAAGGG 360
    |||
QY 1842 CGTCTATCCGGGGGCGCTCGGATGTTGCTCGCCTCAGCGCGCGCGACCTCGACATCT 1901
    |||
DB 361 AATATATCTCGGGGTGATCGGGTTCTTCTGTACACCGCATGTTGATGTAATCTG 420
    |||
QY 1902 CATCCGACCATCGTGTGTCGCGACGCGCGAGGATTCGCGCTGCGCGGGCGATCTG 1961
    |||
DB 421 GATCAGAGACGATGATGTTGTCACGACGAGGTGCGCTGCGAGCAGCGGGCGATTTG 480
    |||
QY 1962 GTCCCTTCCGACCAAGAGAGGATTCACCGAGACCGTGTGTAAGGC 2009
    |||
DB 481 GGGCTGTGTACACCCGAGGCGAGATGACCGCGAGATGATCTCAAGGC 528
    |||
```

RESULT 2 B2570522 845 bp DNA linear GSS 17-DEC-2002

LOCUS B2570522 msh2.1411.xl msh Pseudomonas aeruginosa genomic clone msh2_1411, genomic survey sequence.

ACCESSION B2570522

VERSION B2570522.1 GI:27205583

KEYWORDS GSS.

SOURCE Pseudomonas aeruginosa

ORGANISM Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 845)

REFERENCE 1 Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,

Burns,J.L., Kaul,R. and Olsen,M.V.

Whole-Genome-Sequence Variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

CONTACT: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

FEATURES 1..845

/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="MSH"
/db_xref="taxon:287"
/clone="msh2_1411"
/clone_11b="msh"
/note="Environmental isolate. Whole genomic shotgun library."

ORIGIN

Query Match 10.3%; Score 212.8; DB 8; Length 845;

Best Local Similarity 60.6%; Pred. No. 1.6e-26;

Matches 349; Conservative 0; Mismatches 227; Indels 0; Gaps 0;

```
QY 1486 TCCAACTCCATCAAGAGGACCCGCCCCGGGGGGGACACCGGAGAGAGACGAGCCTC 1545
    |||
DB 69 TCGAATTCCTGCAACCCCAACCCGCCCCGGGCAAGACCCCGAGAGACATGGGCTTG 128
    |||
QY 1546 CCGCGCACTGAGCCGCGCGGAGAGAGACCGGGCCGAGAACTGATGATCTGCACTG 1605
    |||
DB 129 GCGGCGTCTGCTGTGCGCAAGCCCAAGAGACCGGGGAGAAACCTGATGATGCTGAC 188
    |||
QY 1606 GTCCGCAAGACTCAACAGCGTTCGCGATGCTCCGTCCAGCTGCCCGGCTTTC 1665
    |||
DB 189 CTGCGCAAGACATCGAAGCGAGTCCAACTGGCAGGCTACGCTACCGAGCTGTTC 248
    |||
QY 1666 GAGGTGAGACCTAGCGCGCCGCGACCAAGCTGTGTGACCATCCGGGGAGCGGCTGG 1725
    |||
DB 249 GCCCTGAAAGCTATCCAAAGTGCATCACTGTGTAGAGGTCAACCGCAATGGCG 308
    |||
QY 1726 CCGGACAGAGACCGCGCTGCGTACGCGCGCCTTCCGCGCGCTTCATGACCGCG 1785
    |||
DB 309 CCGGGAGAGAGCGCCTCGACCTGCTGGAAGAGAGTTCCCGGGGCGCTGATACCGCG 368
    |||
QY 1786 GCGCCCAAGAGGACCATGAGATCATGACCGCCTTGAGAGAGAGGCCCGGGGCGT 1845
    |||
DB 369 GCGCCCAAGATTCGCCCATGAGATCATGAGATCAAGAACTGGAACCGAGCGGCGATC 428
    |||
QY 1846 TACTCGGGGCGCTGGAATGTTGCGCCACAGCGGCGCGCGACCTCGACATCTGTCATC 1905
    |||
DB 429 TACTGCGGACGCTGTTTCTTCACTCGACGTGCGGGGAGATGACATGCTGATCGCAT 488
    |||
QY 1906 CGCACCATGATGCTGCGCGACGCGCGAGGATTCGCGCTGCGCGGGCGATCTGCTGC 1965
    |||
DB 489 CGCACCTGCTGCTGTCAGAACCGGCGAGTCAATGCTGTGGGGGCGGCGGATCGTCCG 548
    |||
QY 1966 CTCTCCGACCAAGAGAGGATTCACCGAGACCGTGTGTAAGGCCCGCCCATGTGTCAC 2025
    |||
DB 549 GACTCCCACTGGGAGAGACGATCAAGAAACCTGGACACAGGCTGCGGTGCTGTGAA 608
    |||
QY 2026 GCCCTGAGAGGACGCGCGTGGCGGGCGCCCGATGA 2061
    |||
DB 609 ACCCTGAAAGGAGAGGCGCGGACAGGCTTCCCGGA 644
    |||
```

RESULT 3 CF863831

LOCUS CF863831

DEFINITION ps2S008xG04f USDA-IFAPs: Expression of Phytophthora sojae genes

during infection and propagation_s2S Phytophthora sojae cDNA clone

s2S008G04 5, mRNA sequence.

ACCESSION CF863831

VERSION CF863831.1 GI:38118457

KEYWORDS EST.

SOURCE Phytophthora sojae

ORGANISM Phytophthora sojae

Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae;

Phytophthora.

1 (bases 1 to 653)

REFERENCE 1 Tyler,B. Not Published

Unpublished (2003)

CONTACT: Tyler B

Tyler lab

VBI
1880 Pratt Dr., Blackeburg, VA 24061, USA
Tel: 540-231-7318
Email: bmttyler@vt.edu
PCR Primers

FORWARD: BK reverse primer
BACKWARD: BK reverse primer
Plate: 008 row: G column: 04
Seq primer: BK reverse primer
High quality sequence stop: 653.
Location/Qualifiers

FEATURES

source

1. 653
/organism="Phytophthora sojae"
/mol_type="mRNA"
/db_xref="taxon:67593"
/clone="BZ508604"
/issue_type="Zoozoores"
/cell_line="P6497"
/dev_stage="Free swimming"
/clone_lib="USDA-IPAFS: Expression of Phytophthora sojae genes during infection and propagation, BZ5"
/note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match 10.0%; Score 205.8; DB 7; Length 653;
Best Local Similarity 58.3%; Pred. No. 2.6e-25;

Matches 379; Conservative 0; Mismatches 268; Indels 3; Gaps 1;

1280 AGGAGATCCGCAACGGGAGTCTGACGAGATCTGCTGACCAATGTCACCGCGCGA 1339
4 ACCAGGACGGCGACATCTTCCAGGCGGTGCGCTCCAGCGGCTGCTGAGACTCCCGA 63
1340 CCGAGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1399
64 AGGACGTCACCGGCTGATCTGATCCGCAAGATGCGCGCATCAACCTGCTCCGCTACA 123
1400 GCGCCCTGCTGCA---GTCCCGCACTGTCGAGTCTGAGCGCTGCGCGCGCGGCTCC 1456
124 TGTCTTCTTCTGACATGAGGAGAGACTTCCAGATGTCGGCGCTTCCCGAATGCTCG 183
1457 TCAAGATCGGCGCGCGCGCTGAGTCTCAAGCCCATCAAGGGAGACCGCGCGCGG 1516
184 TCAAGGTGACATGACCGTGTGTGAGAGACCGCATTTGCGCGTGAACGCGCACCTG 243
1517 GCGGACCGCGGAGAGAGAGAGCGGCTCGCGCGCACTGCGCGCGCGGAGAGAGACC 1576
244 GCGCCATGACGAGAGAGAGAGAGCACTGTGAAGGACTGCTGCGCGAGAGAGAGAGC 303
1577 GGGCCGAGAACCTGATGATGTCGACCTGCTCGGCAACGACTCAACAGGCTCTGGCGCA 1636
304 GCGCCGAGACATCATGCTGTGTGAGCTGTGGCGGCAAGAGCTGGCGCGGCGCAAGC 363
1637 TCGGCTCTGTCACATGTCCTCGGCTCTTCAAGGTGAGACTTACGCGCGCTGACACG 1696
364 CCGGATCTGCTCGGAGTGAAGCGCTCATGCAATTTGAAGATCTCGCAAGTATGACAC 423
1697 TGGTGTGACCATTCGGGAGAGCGGCTGCGCGCGCAACAGACCGCGCTGTGCTAGCG 1756
424 TCTGTGCTGTCGTCGAGGCGGACCTGTGCGAGAGACCGGACCGTGTGACAGCTTACG 483
1757 CGGCTTCTCCCGCGGCTCATGACCGGCGCGCGCGCAAGAGCGCAACATGAGATCATG 1816
484 CATGTTCTCTGCGGAGCGCTGTCCGCGCGCTTAAAGTGTGTGAGATGAGACTCATCT 543
1817 ACCGCTGAGAGAGCGCGCGGCGGCTTACTTCCGCGCGCTCGGATGTTGCGCTTCA 1876
544 GCTGCTGAGAGAGAGAGAGCGCGGCTGTACTTCCGCGCTCGGCTTCAAGTTTCT 603
1877 GCGGCGCGCGGACCTTACGATGTCATTCGCAACCATGTCGTCGCGGAC 1926
604 CCGGCTTCTGAGACAGCGGATCGCATCGGACCATGTGTGTCAAGAGAC 653

RESULT 4

BZ548859

LOCUS

BZ548859 1043 bp DNA linear GSS 17-DEC-2002
pacsl-60 1509, sl pacsl-60 Pseudomonas aeruginosa genomic clone

DEFINITION

pacsl-60 1509, genomic survey sequence.

ACCESSION

BZ548859

BZ548859.1 GI:27152440

VERSION

GSS.

KEYWORDS

Pseudomonas aeruginosa

SOURCE

Pseudomonas aeruginosa

ORGANISM

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

REFERENCE

1 (bases 1 to 1043)
Pseudomonadaceae; Pseudomonas.

AUTHORS

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

TITLE

Burns, J.L., Kaul, R. and Olsen, M.V.

JOURNAL

Whole-Genome-Sequence variation among multiple isolates of

COMMENT

Pseudomonas aeruginosa library

J. Bacteriol.

(2002) In press

CONTACT

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145,

Seattle, WA 98105-2145, USA

Tel: 2062216954

Fax: 2066857244

Email: ckraymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

1. 1043

/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"

/strain="1-60"

/db_xref="taxon:287"

/clone_lib="pacsl-60 1509"

/note="clinical isolate 1-60 Whole genomic shotgun library."

ORIGIN

Query Match

9.5%; Score 195.6; DB 8; Length 1043;

Best Local Similarity

62.4%; Pred. No. 1.4e-23;

Matches 306; Conservative 0; Mismatches 184; Indels 0; Gaps 0;

1519 GGCACCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1578
61 GCACCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
1579 GCGGAGAACCTGATGATGTCGACCTGTCGAGTCTGCAACGACCTCAACAGCGCTGCGGATC 1638
121 GCGGAGAACCTGATGATGTCGACCTGTCGAGTCTGCAACGACCTCAACAGCGCTGCGGATC 180
1639 GGCCTGCTGTCACATGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1698
181 GCGAGGTGACGAGTACCGGAGCTGTTGCGCGCGGAGAGAGTATCCCAAGCTGATACCTG 240
1699 GTGTGACCATTCGGGAGAGCGCTGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1758
241 GTGAGAGAGTACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 300
1759 GCGTTCCTCCCGGAGGCTCATGACCGGCGCGCGCGCAAGAGCGCAACATGAGATCATGAC 1818
301 AGCTTCTCCCGGAGGCTCATGACCGGCGCGCGCGCAAGAGATTCGCGCAATGAGATCATGAC 360
1819 CGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1878
361 GAGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
1879 GCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1938
421 GCGGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
1939 TTGCGGCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1998

Db 481 TCGTGGGCGCGCGGCATGTCGCCGACTCGCACTGGGAGACGATACGAAACC 540
QY 1999 GTGTAAAG 2008
Db 541 CTGGACAAAG 550

RESULT 5
BZ412600/c
LOCUS BZ412600 ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0015004,
DEFINITION genomic survey sequence.
ACCESSION BZ412600
VERSION BZ412600.1 GI:26047761
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 523)
WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSs: OGAAG11TC
Contact: Cathy Whitelaw
TIGR

TITLE 9712 Medical Center Drive, Rockville, MD 20850, USA
JOURNAL Tel: 301-838-5843
COMMENT Fax: 301-838-0208
Email: whitelaw@tigr.org
Seg primer: TR
Class: sheared ends.

FEATURES
Location/Qualifiers
source 1..523
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBma0015004"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 8.6%; Score 178; DB 8; Length 523;
Best Local Similarity 61.8%; Pred. No. 1.4e-20;
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1559 CCGGCCGAGAGAGACCGGCGCGAGAACTGATCGGACCTGTCGCAAGACC 1618
Db 509 CCACACACGAAAGACCGAGCGAGAACTGATGATGTCGACCTTTAAAGAAAGATC 450

QY 1619 TCAACAGCGTCTGGCGCATCGCTCCGTCCAGCTGCCCGCTCTTGAAGTGAAGACT 1678
Db 449 TCGGCAAGGTCTGGAGCCAGGGAGCGTGATGCCAGCGCTCATGGAAGTTCAT 390

QY 1679 ACCGCGCGGTGCACCACTGGTGCACATTCGGGGAGCGCTGGCGCCGCGACCA 1738
Db 389 ACAGAAAGTGTCAACCATGATGATGATCGCTGCGGAAACAAAGACGATCTAAGCC 330

QY 1739 CCGCGCGCTGGTACGCGCGCGCTCCCGCGCGCTCCATGACCGCGCGCGCAAGAGC 1798
Db 329 CCGTGTATGTGTAAAGCGGCTTTTCAGCGGCTTCATGATGACCGGCGCTCCAAAGTCA 270

QY 1799 GCACCATGAGATCATCGACCGCTGGAGAGAGCGCCCGGCGGTCTACTTCGGGCGC 1858
Db 269 GATCAATGAGATCTTGATTCATCTGATGACCAAGCCCAAGAGATTAATTCAGAGATCA 210

QY 1859 TCGGATGTTTCGCGCTGAGCGCGCGCGACCTCAGCATGTCATCGGACCATCGTGC 1918

Db 209 TCGGCTTCTTCTGCAACCAACACATTCGACCTGAACATTGTATCCGACTGTTATCC 150
QY 1919 TGGCCGACCGCCAGCGAGAGTTCGGCGTCCGCGGCGCATCTGTCCTCTCCGACGAG 1978
Db 149 TGCACGACGGAAGAGCTCGTCCGGGCGGGGCGATCTGTACGCGCTGTGAGCCCGAG 90

QY 1979 AGAGAGATTCAACCGAGACCGTGTAAAGCCCGCGCC 2016
Db 89 AAGCAGAGTACGACGAAATGTGCTGAAGCGCAAGGCC 52

RESULT 6
BZ412593
LOCUS BZ412593
DEFINITION OGAAG11TC ZM_0.7_1.5_KB Zea mays genomic clone ZMMBma0015004,
genomic survey sequence.
ACCESSION BZ412593
VERSION BZ412593.1 GI:26047739
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 631)
WhiteLaw,C.A., Quackenbush,J., Van Aken,S., Uterback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSs: OGAAG11TC
Contact: Cathy Whitelaw
TIGR

TITLE 9712 Medical Center Drive, Rockville, MD 20850, USA
JOURNAL Tel: 301-838-5843
COMMENT Fax: 301-838-0208
Email: whitelaw@tigr.org
Seg primer: TP
Class: sheared ends.

FEATURES
Location/Qualifiers
source 1..631
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone_lib="ZMMBma0015004"
/note="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN
Query Match 8.6%; Score 178; DB 8; Length 631;
Best Local Similarity 61.8%; Pred. No. 1.4e-20;
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1559 CCGGCCGAGAGAGACCGGCGCGAGAACTGATGATCGTCCGACCTGTCGCAAGACC 1618
Db 15 CCACACGAGAAAGACCGAGCGCGAGAACTGATGATGTCGACCTTTAAAGAAAGATC 74

QY 1619 TCAACAGCGTCTGGCGCATCGCTCCGTCCAGCTGCCCGCTCTTGAAGTGAAGACT 1678
Db 75 TCGGCAAGGTCTGGAGCCAGGGAGCGTGATGCCAGCGCTCATGGAAGTTCAT 134

QY 1679 ACCGCGCGGTGCACCACTGGTGCACATTCGGGGAGCGCTGGCGCCGCGACCA 1738
Db 135 ACAGAAAGTGTCAACCATGATGATGATCGCTGCGGAAACAAAGACGATCTTAAGCC 194

QY 1739 CCGCGCGCTGGTACGCGCGCGCTCCCGCGCGCTCCATGACCGGCGCGCCAAAGAGC 1798
Db 195 CCGTGTATGTGTAAAGCGGCTTTTCAGCGGCTTCATGATGACCGGCGCTCCAAAGTCA 254

QY 1799 GCACCATGAGATCATCGACCGCTGGAGAGAGCGCCCGGCGGTCTACTTCGGGCGC 1858

Db 255 GATCAATGAGATCTTGAATTCATTGAGACAAAGCCAGAGAGATGTAAGATCGA 314

Qy 1859 TCGGATGTTTCGCGCTCAGCGGCGCCGACCTCAGCATGTCATCCGACCATCGTCG 1918

Db 315 TCGGATGTTTCGCTGTAACAACACATTCGACCTGAAACATTTGATCCGAACTGTATCC 374

Qy 1919 TGGCCGACGGCCAGCGAGAGTTTCGCGCTCGCGCGGGCGATGTCCTCCCTCCGACGAG 1978

Db 375 TGCACGACGGAGAAAGCTCGGTCGGGGCGGGGTGGGGCGATGTAGCCGTCTCGAGCCAG 434

Qy 1979 AGGAGAGTTTACCCGAGACCGTGTAAAGCCCGCCG 2016

Db 435 AAGCAGAGTACGACGAATGCTCTGAAAGCGAAAGCC 472

RESULT 7
LOCUS B537360 631 bp DNA linear GSS 16-DEC-2002
DEFINITION OGAGD95TC ZM2.0.7.1.5 KB Zea mays genomic clone ZM8BMA0053P22,
genomic survey sequence.
ACCESSION B537360
VERSION B537360.1 GI:27085796
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 631)
AUTHORS WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uteerback,T.,
Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
Clek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
CONSORTIUM for Maize Genomics
Unpublished (2002)
COMMENT Contact: Cathy WhiteIaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whiteIaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
source location/Qualifiers
1..631
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM8BMA0053P22"
/note="Vector: pBCK-; Site_1: HincII; 0.7-1.5 kb
methylation filtered genomic DNA library"

ORIGIN

Query Match 8.6%; Score 178; DB 8; Length 631;
Best Local Similarity 61.8%; Pred. No. 1.4e-20;
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 1559 CCGGCGGAGAGAGACCGGCGGAGAACTGATGATCGTCACTGTGTCGCGACGACC 1618

Db 15 CACACGACGAAAAAGACGAGCGGAGAACTGATGATGTCGACCTCTTAAGAAACATC 74

Qy 1619 TCAACAGCGTCTGCGGCGATCGCTCCGTCCACGTGCCCCGCTCTTTCAGAGTGAAGACT 1678

Db 75 TCGGCAAGGTCTGCGAGCCGAGGAGCTGATGTCACCGCTCATGAGAGTTGAATCAT 134

Qy 1679 ACGGCGCCGTCACACGATGATGTCACATCCGCGGAGCGGCTGCGGCGCGACGACGA 1738

Db 135 ACAGAACTGTTACACCATGATGATGACGCTGCGGAGACAAAGAACAGATCTAAGCC 194

Qy 1739 CCGCGCGCTGCGTACGCGCGCTTCCCGCGGCGCTCCATGACGCGCGCCCAAGAGC 1798

Db 195 CCGTGTATGTGTGAAGAGCGGCTTTCAGAGCGGTTGATGACCGGCGCTCCAAAGTCA 254

Qy 1799 GCAACATGAGATCATGACCGGCTGAGAGAACCCCGGGCGTCTACTCCGGGCGC 1858

Db 255 GATCAATGAGATCTTGAATTCATTGAGACAAAGCCAGAGAGATGTAAGATCGA 314

Qy 1859 TCGGATGTTTCGCGCTCAGCGGCGCCGACCTCAGCATGTCATCCGACCATCGTCG 1918

Db 315 TCGGATGTTTCGCTGTAACAACACATTCGACCTGAAACATTTGATCCGAACTGTATCC 374

Qy 1919 TGGCCGACGGCCAGCGAGAGTTTCGCGCTCGCGCGGGCGATGTCCTCCCTCCGACGAG 1978

Db 375 TGCACGACGGAGAAAGCTCGGTCGGGGCGGGGTGGGGCGATGTAGCCGTCTCGAGCCAG 434

Qy 1979 AGGAGAGTTTACCCGAGACCGTGTAAAGCCCGCCG 2016

Db 435 AAGCAGAGTACGACGAATGCTCTGAAAGCGAAAGCC 472

RESULT 8
LOCUS CG170785/c 810 bp DNA linear GSS 21-AUG-2003
DEFINITION PU1R58TD ZM_0.6.1.0 KB Zea mays genomic clone ZMBTRA0592120,
genomic survey sequence.
ACCESSION CG170785
VERSION CG170785.1 GI:34061583
KEYWORDS GSS.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 810)
AUTHORS WhiteIaw,C.A., Quackenbush,J., Van Aken,S., Uteerback,T.,
Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and
Bennetzen,J.
CONSORTIUM Maize Genomics Consortium
Unpublished (2003)
COMMENT Other_GSSs: PU1R58TB
Contact: Cathy WhiteIaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whiteIaw@tigr.org
Seq primer: TF
Class: sheared ends.
FEATURES
source location/Qualifiers
1..810
/organism="Zea mays"
/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM8BTA0592120"
/note="Vector: pCR4-TOPO; Site_1: EcoRI; 0.6-1.0 kb high
Cor selected genomic DNA library"

ORIGIN

Query Match 8.6%; Score 178; DB 9; Length 810;
Best Local Similarity 61.8%; Pred. No. 1.4e-20;
Matches 283; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

Qy 1559 CCGGCGGAGAGAGACCGGCGGAGAACTGATGATCGTCACTGTGTCGCGACGACC 1618

Db 663 CACACGACGAAAAAGACGAGCGGAGAACTGATGATGTCGACCTCTTAAGAAACATC 604

Qy 1619 TCAACAGCGTCTGCGGCGATCGCTCCGTCCACGTGCCCCGCTCTTTCAGAGTGAAGACT 1678

Db 603 TCGGCAAGGTCTGCGAGCCGAGGAGCGTATGTCACCGCTCATGAGAGTTGAATCAT 544

Qy 1679 ACGGCGCCGTCACACGATGATGTCACATCCGCGGAGCGGCTGCGGCGCGACGACGA 1738

Db 543 ACAGAACTGTTACACCATGATGATGACGCTGCGGAGACAAAGAACAGATCTAAGCC 484

QY 1739 CCGCCGCGCTGCGTACGCGCGCCCTTCCCGCGGCTTCATGACCGCGCGCCCAAGAGC 1798
|||
Db 483 CCGTGTATGTGTGAAGAGCGGCGCTTTCAGGCGGTTGATGACCGGGGCTCCAAAGTCA 424
|||
QY 1799 GCACCATGAGATCATGACCGCGCTGAGAGAGCCCGCGGCGTCTACTCCGGGGCGC 1858
|||
Db 423 GATCAATGAGATCTTGTATTCATTGAGCAAGCCCAAGAGAGTGTACTCTCAGATTCGA 364
|||
QY 1859 TCGGATGTGTCGCGCTCAGCGCGCGCGCGCGCATGCTCATGCTCCGACCATGCTG 1918
|||
Db 363 TCGGGTCTTCTGTATCAACACACATTGCACTGAAACATTTGATTCGAACTTTATCC 304
|||
QY 1919 TGGCCGCGCGCAGCGCGGAGTTCCGCGCTCGCGGGCGATGCTGCTCTTCCGACGAG 1978
|||
Db 303 TGCACGACGGAAGAGCTCGGTGCGGGCGGGGATGATGATGAGCGCTGCGAGCCGAG 244
|||
QY 1979 AGGAGAGTTACCGAGACCGGTGTAAAGCCCGCGCC 2016
|||
Db 243 AAGCAGAGTACGACGAATGCTGTGAAAGCGAAAGCC 206
|||

RESULT 9
CF598037 579 bp mRNA linear EST 26-SEP-2003
LOCUS
DEFINITION NC_053906.1.1 NC-LIV Tachyzoite cDNA library Neospora caninum
ACCESSION
VERSION CF598037
KEYWORDS CF598037.1 GI:36356095
EST.

SOURCE
ORGANISM Neospora caninum
Neospora caninum
Eukaryote; Alveolata; Apicomplexa; Coccidia; Eimeriida;
Sarcocystidae; Neospora.

REFERENCE
AUTHORS 1 (bases 1 to 579)
Cole, R., Fogarty, S., Tang, K., Howe, D. K., Sibley, L. D., Clifton, S.,
Warr, M., Hillier, L., Pape, D., Martin, J., Wylie, T., Theising, B.,
Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Ronko, I.,
Teagarden, V., R., Fedele, M., Belaygorod, L., Franklin, C.,
Carr, L. M., Grow, A., Maguire, L., Wadkins, J., Ritchey, J., Waterston, R.
and Wilson, R.

TITLE
JOURNAL
COMMENT
USDA-WashU Neospora EST Project
Unpublished (2000)
Contact: Sandy Clifton, Ph.D. - Neospora
USDA-WashU Neospora EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Contact David Sibley (toxoe@borcim.wustl.edu) for further
information relating to organism, libraries, or clone availability.
Seq primer: -40RP from Gibco.

FEATURES
source
1. 579
Location/Qualifiers

/organism="Neospora caninum"
/mol_type="mRNA"
/db_xref="taxon:29176"
/dev_stage="Tachyzoite"
/lab_host="Electron Blue cells (Stratagene)"
/clone_lib="NC-LIV Tachyzoite cDNA library"
/note="Vector: pBluescript II SK+ vector type: plasmid;
Site 1: EcoRI; Site 2: XhoI; The cDNA library was
constructed by Kelian Tang, and Robert Cole at Washington
University. cDNA was synthesized from Poly(A)+ mRNA using
an oligo-d(T) primer containing a XhoI site. Following
second strand synthesis, EcoRI adapters were ligated to
the cDNA, and products were size-selected on Sephadex
5500. The cDNA were directionally cloned into the
EcoRI/XhoI prepared pBluescript II SK+ vector, and
electroporated into Electron Blue cells (Stratagene).
The library may contain a small percentage of host or
bacterial contaminants."

ORIGIN

Query Match 8.5%; Score 174.2; DB 7; Length 579;
Best Local Similarity 61.5%; Pred. No. 6, 4e-20;
Matches 297; Conservative 0; Mismatches 183; Indels 3; Gaps 1;

QY 1444 CCGAGCGGTTCTTCAGATCGCGCGCGCGCGGCTGAGTCCAGCCATCAAGGGG 1503
|||
Db 76 CCGAAGCGCTTCTTAATGAAAGATGAGCGGTTGGCTGAGTCCGAACCATCAAGGGT 135
|||
QY 1504 ACCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1563
|||
Db 136 ACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 195
|||
QY 1564 CCGAGAGAGAGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1623
|||
Db 196 AAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 255
|||
QY 1624 ACGGTCTGCGCGATCGCGCTCGCGTCCAGCGTCCCGCGCTTTCAGGTGAGACCTA 1683
|||
Db 256 CGGTCTGTATTCAGAGAACCGTGCACGTTCTGAGTTGATGATGACAGAGACCTA 315
|||
QY 1684 CCGGTGACACAGCTGTGTGACACATCCGCGGAGCGGCTGCGG---CCGCGACACAGC 1740
|||
Db 316 AAGGTGACATCACTGTGTGACCAATCCGAGCGAGCTGAGGAAACCGAAGTGAATTT 375
|||
QY 1741 GCGGCTGCGGTACGCGCGCGCTTCCCGCGGCTCCATGACCGCGCGCGCGCGCG 1800
|||
Db 376 CTTGACGCTGTATATTCGCGCATTTCCAGAGGCTCGATACAGGCGCTCCGAGAGCG 435
|||
QY 1801 ACCATGAGATCATGACCGCGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
|||
Db 436 GCAATGAGATCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 495
|||
QY 1861 GAGTGTGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1920
|||
Db 496 GCGTCTGCTGTGCGTACAGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 555
|||
QY 1921 GCC 1923
|||
Db 556 ACC 558
|||

RESULT 10
B2892725/c 660 bp DNA linear GSS 30-JUL-2003
LOCUS
DEFINITION HM7_0152.X1 059.ab1 Hm pUC18 library Haloarcula marismortui ATCC
43049 genomic 5', genomic survey sequence.
ACCESSION
VERSION B2892725
KEYWORDS B2892725.1 GI:33343358
GSS.

SOURCE
ORGANISM Haloarcula marismortui ATCC 43049
Haloarcula marismortui ATCC 43049
Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.

REFERENCE
AUTHORS 1 (bases 1 to 660)
Goos, Y., Roach, J., Glusman, G., Baliga, N. S., Deutsch, K., Pan, M.,
Dasgupta, S., Ng, W. V., and Hood, L.

TITLE
JOURNAL
COMMENT
Low-pass Sequencing for Microbial Comparative Genomics
Unpublished (2003)
Contact: Goos Y
Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygoos@systemsbiology.org
Seq primer: M13 Forward
Class: shotgun.

FEATURES
source
1. 660
Location/Qualifiers
/organism="Haloarcula marismortui ATCC 43049"
/mol_type="genomic DNA"
/strain="ATCC 43049"

/db_xref="taxon:272569"
/clone_1lb="Hm pUC18 Library"
/note="Vector: pUC18; Site 1: SmaI; A shotgun library was
constructed from Haloarcula marismortui genomic DNA using
pUC18/SmaI/BAP plasmid"

ORIGIN

Query Match 8.3%; Score 171.8; DB 8; Length 660;
Best Local Similarity 56.1%; Pred. No. 1,66-19;
Matches 345; Conservative 0; Mismatches 267; Indels 3; Gaps 1;

QY 1239 CGACAAAGAGCGCTTACCTCAAGCGCATCGAGCTGCTCAAGAGATCCGCAACGGCGA 1298
DB 619 CGGGAAGACGCGTTCGCCGCGCGCTCGGCAAAATAAACAAGTACGTACAGGAGCGCGA 560
QY 1299 GTGTGATGAGATCTGCTGACCAACATGTCACCGCCCGACCGAGCGACCGCTTCC 1358
DB 559 CAGGTTCCACAAACCTCTCACACCGCTGACCGCCCGCGCGCTCCACCCAGTCA 500
QY 1359 GCTTACTCCGCGCGCGCATCAGCCCGCTGCTGAGGCGCGCTGCTGAGTTCC 1418
DB 499 CACCTTGAAGCTGTCTGCGGCTGAAATCCGCTCTGACTCAGCCCTGCTGAGTTCC 440
QY 1419 CGAAGTGTGCTGCTGACCGCTCGCCGAGCGGTTCTCAAGATCGGCGCGACGGCG 1478
DB 439 CGGCGTGAACCTTGTACGCGCGACCGCGAGCTGTTCT---GGACGTGATGGCGACA 383
QY 1479 GGTGAGTCCAAAGCCATCAAGGAGACCGCCCGCGCGGCGACCGCGAGAGACGA 1538
DB 382 GTTGTCTACGAGCGGATGCTCCCGACCGCGCGCGCGCGACGCTGTCGAAAGCA 323
QY 1539 GCGGCTCCGCGCGACCTGCGCGCGCGGAGAAAGGACCGGCGGAGAACTGATGATCT 1598
DB 322 GGAAGCTTGAAGTGAAGCTTCTGTACGACGAGAGAGAGCGGCTGACACCCCATGCTGT 263
QY 1599 CGAAGTGTGCTGACCAAGACCTCAACAGCTGCTGCGAGTGGCTCCGCTCCACGTGCGCG 1658
DB 262 CGAAGTGAAGCGCAATGACCTCGGCAAGTCAAGAGATAGGTCCTGACAGTGGCGCA 203
QY 1659 GCTTCTGAGGTGAGACCTACCGCGCGCTGACCAAGCTGCTGACCAATCCGAGGAG 1718
DB 202 GTACCGCGCGCTGACCGGATTTCCGAAAGAAAGCACTGTTTCCCTCAATCGAAGGGA 143
QY 1719 GCTGCGCGCGCGGACCAAGACCGCGCGCTGCTGACGCGCGCTTCCCGCGGCTTCAT 1778
DB 142 ATTACGAGACGCGGTGACATCGCGACGCGGTGACGCGTGTCTCTGTGAGACAT 83
QY 1779 GACCGCGCGCGCAAGAGCGCACATGAGATCATGACCGCTGAGAGAGAGCGCGCG 1838
DB 82 CACCGCGCGCGCAAGCGCGGAGCATGAGATTATGACAGAGTGAAGCGGACCGGACG 23
QY 1839 GGGCGTCTACTCGG 1853
DB 22 GGGCGCTTACACCG 8

RESULT 11
C0024641/c 894 bp mRNA linear EST 10-JUN-2004
LOCUS C0024641
DEFINITION kb Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
ACCESSION C0024641
VERSION C0024641.1 GI:48551893
KEYWORDS EST.
SOURCE Coccidioides posadasii
ORGANISM Coccidioides posadasii
Eukaryote; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Omygenales; mitosporic Omygenales; Coccidioides.
REFERENCE 1 (bases 1 to 894)
AUTHORS Gardner,M.J. and Cole,G.T.
TITLE Analysis of gene expression in Coccidioides posadasii mycelia and
spherules via expressed sequence tags
JOURNAL Unpublished (2003)

COMMENT

Other ESTs: EST803026
Contact: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@igr.org.

FEATURES
source Location/Qualifiers

1..894
/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="C1FA227"
/dev_stage="spherules"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_1lb="Coccidioides posadasii spherule cDNA library,
0.5 to 5.3 kb"
/note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
kb"

ORIGIN

Query Match 8.1%; Score 167.8; DB 7; Length 894;
Best Local Similarity 52.6%; Pred. No. 7,66-19;
Matches 442; Conservative 0; Mismatches 387; Indels 12; Gaps 3;

QY 1231 GCGCGCACGACAAAGAGCGCTTACCTCAAGCGCATGACGAGTGCCTCAAGAGATCCG 1290
DB 841 GCCCTGAGACGACGAACATGTCAGCCAGAGATCGGCGCGCACAGAAATGCCATTAA 782
QY 1291 AAGGAGTGTCTGACAGATCTGCTGACCAATGATGATGATGATGATGATGATGAT 1347
DB 781 GAGGCGGAGCGGTATAGGTGACCAATGACGACAAATTAAGGATGATGATGATGAT 722
QY 1348 ACGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1407
DB 721 GACCCATAGCATATATATCTCTCTCTGAGAAACGTAACCCCGCTTACTAGGATAT 662
QY 1408 CTCGAGTTCC-----CGAAGTGTGCTGAGGCGCTCGCGCGGCTTCTTACG 1461
DB 661 ATGAGCTTCGCTGCTCAACAGACCAACCATATCTTCTGCGCGGAGGTTATATTC 602
QY 1462 ATGAGGCGCGAGCGGCGGCTGAGTCAAGTCCATCAAGGAGACCGCGCGCGCGG 1521
DB 601 ATGATGCTGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 542
QY 1522 ACCGCGGAGAGAGAGAGGCGGCTCGCGCGGCTGCTGCGCGGCGGAGAGAGAG 1581
DB 541 GACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
QY 1582 GAGAACCTGATGATCTGACCTGCTGCGCAAGACCTGACAGCTGCTGCGGATCGG 1641
DB 481 GAGAACTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 422
QY 1642 TCCGTCAAGTGTGCGCGGCTTTCAGAGTGAAGACTTACGCGCGCTGCAACAGCTG 1701
DB 421 TCAATCAAGTGTGCGCGGCTTTCAGAGTGAAGACTTACGCGCGCTGCAACAGCTG 362
QY 1702 TCGACCATCGGAGAGAGGCTGCGCGGCTGCGCGGCTGCGCGGCTGCGCGGCTG 1761
DB 361 ACAACATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCATTCAT 302
QY 1762 TTCGCGCGGCTGCTCAATGACGCGCGGCGCAAGAGCGGACCATGAGATGATGAC 1821
DB 301 TTCGCGCGGCTGCTCAATGACGCGCGGCGCAAGAGCGGACCATGAGATGATGAC 242
QY 1822 CTGAGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1878
DB 241 CTGAGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 182
QY 1879 GCGCGCGCGAGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTGATGCTG 1938

Db 181 GGCACCGTCGATCAGTCGGTGGTTATCCGATCAATGTAAAGTACGGAAGACAGCTTGA 122

QY 1939 TTGGGCGGTGGCGGGGGGATCGTGTCCCTCTCCGACGAGGAGAGATTCAACGAGACC 1998

Db 121 CTAGGTCTGAGAGGGGCGATTAATTGCTCAGCGAGGCGAGAAAGATGGAGTGAAGTC 62

QY 1999 GTGTAAAGGCGCGCCGATGTCAACCGCTCGACGCGACGCGCGTGGCGGGCGCCGA 2058

Db 61 ATGCTGAAGCGGATGTGTTGCTACGCGCGCTCGCTAGGAATCGGGCGCAGATGCTGA 2

QY 2059 T 2059

Db 1 T 1

RESULT 12

BZ896109 757 bp DNA linear GSS 30-JUL-2003

LOCUS BZ896109

DEFINITION NARPE_0132 Na pUC18 library Natrialba asiatica genomic 5', genomic survey sequence.

ACCESSION BZ896109

VERSION BZ896109.1 GI:33346585

KEYWORDS GSS.

SOURCE Natrialba asiatica

ORGANISM Natrialba asiatica

REFERENCE 1 (bases 1 to 757) Go, Y., Koach, U., Glusman, G., Balliga, N. S., Deutsch, K., Pan, M., Dassa, S., Ng, W. V., and Hood, L. Low-pass Sequencing for Microbial Comparative Genomics Unpublished (2003)

TITLE JOURNAL

COMMENT Institute for Systems Biology 1441 North 34th Street, Seattle, WA 98103, USA Tel: 206 732 1412 Fax: 206 732 1299 Email: YGo@systemsbiology.org Seg primer: M13 Forward Class: shotgun.

FEATURES

source 1..757

/organism="Natrialba asiatica"

/mol_type="genomic DNA"

/strain="ATCC 700177"

/db_xref="taxon:64602"

/clone_lib="Na pUC18 Library"

/note="Vector: pUC18, Site 1: SmaI; A shotgun library was constructed from Natrialba asiatica genomic DNA using pUC18/SmaI/BAP plasmid"

ORIGIN

Query Match 8.1%; Score 167.2; DB 8; Length 757; Best Local Similarity 57.9%; Pred. No. 9.8e-19; Matches 334; Conservative 0; Mismatches 239; Indels 4; Gaps 2;

QY 1361 TCTACTCCGGCGCGGCCATCAGCCCCCGTCCGTAGCGCGCCCTGCTGATTCCTCCCG 1420

Db 90 TCTACGAGGCGATGCGGAGATCAACCCCTGCGCTGATCTGCTGAGACGACG 149

QY 1421 AACTGTGGTGTAGAGCGCTCGCCGAGCGGTTCTCAGATCGCGCGAGCGCGCG 1480

Db 150 ACCGTACCGTGTGTGCGGCAAGCCCCGAAACGCTGCTATGGTGGCGGGC--GAGAG 206

QY 1481 TCGATCCAAGCCCATCAAGGAGACCGCCCCCGGGCGGACCGCGAGAGAGAGCAGC 1540

Db 207 TCATGCGGAATCCGATCGCGGCACCTGCGACCGCGGAGAGAGCCCGTCGAGGACGCG 266

QY 1541 GGCCTCCGCGCGACCTGGCGCGCGCGGAGAAAGACCGGGCGAGAACTGATATGCTG 1600

Db 267 GACTGGCGGGGAGATCTCGCAGACGGGAGAGCGCGCGCAACACCATGCTGTGTCG 326

QY 1601 ACCTGTCCGCAAGCCTCAAGCGCTGCGCGATCGCTCCGTCACGTGCCCCGGC 1660

Db 327 ACCTCGCGGAACCATGTACGCGCGCTCTCGCGCGCGGCTCCGTCCGCTGACGAGT 386

QY 1661 TCTTGGAGTGGAGACCTACGCGCCCGTGCACACACTGGTGTGACCAATCCGGGGACGCG 1720

Db 387 TCATGAACGTGCTGAAGTACACGCACTCAGCAATGAGTGAAGCGGAGCGAGT 446

QY 1721 TCGGCGCGGACCAAGACCGCGCGCTGGTACGCGCGCTTCCCGCGGCTCATGA 1780

Db 447 TGGCGCGGACCGCCGACGCGTTGACGCGAGCGGGCGCTGTTCCCGGGGAGCGTCT 506

QY 1781 CCGGCGCGCCCAAGAAAGCGCACATGAGATATGACCGCTGAGAGAGGCCCGCGG 1840

Db 507 CGGGCGCGCGCAAGATCCGCGCATGAGCTCATGACGAATCGACGACCGACCGCGCG 566

QY 1841 GCGTACTCGGGGGCGC-TGGGATGTTCGCGCTCAGCGCGCGCGGACCTCAGCATC 1899

Db 567 GGCTGTACGCGGAGGCGCTCGCTACTACTCTGAGCGCGCGGATTCGCGATC 626

QY 1900 GTCATCCGACCATCGTGTGCGCGACGCGCGAGCGG 1936

Db 627 GTCATTCGACGCGCACGCTGGAAGCGAGCGCGCG 663

RESULT 13

CO028609 750 bp mRNA linear EST 10-JUN-2004

LOCUS CO028609/c

DEFINITION EST806993 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb Coccidioides posadasii cDNA clone CTFAS82 3' end, mRNA sequence.

ACCESSION CO028609

VERSION CO028609.1 GI:4859837

KEYWORDS EST.

SOURCE Coccidioides posadasii

ORGANISM Coccidioides posadasii

REFERENCE 1 (bases 1 to 750) Onyiah, S., Fungi, Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides. Gardner, M. J., and Cole, G. T. Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags Unpublished (2003)

TITLE JOURNAL

COMMENT Other ESTs: EST806994

Contact: Gardner MJ

The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301 838 3519 Fax: 301 838 0208 Email: gardner@tigr.org.

FEATURES

source 1..750

/organism="Coccidioides posadasii"

/mol_type="mRNA"

/strain="CTFAS"

/db_xref="taxon:199306"

/clone_lib="CTFAS82"

/dev_stage="spherules"

/lab_host="E. coli DH10B, T1 phage resistant"

/clone_lib="Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"

/note="Vector: pEXpress 1; Site 1: Not I; Site 2: Eco RV; Coccidioides posadasii spherule cDNA library, 0.5 to 5.3 kb"

ORIGIN

Query Match 8.0%; Score 164.8; DB 7; Length 750; Best Local Similarity 55.1%; Pred. No. 2.5e-18; Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

QY 1439 CCTCGCGGAGGGTCTTCACAGATCGCGCGCGCGCGCTGACAGTCAAGCCCATTA 1498

Db 624 CGTCGCCGAGAGGTTTATATCATGATGTGATGCGCTGTGCTGAATGAACCAATTA 565

QY 1499 AGGGAGACCGCGCGCGGCGGACCGCGAGAGAGACGAGCGGCTCGCGCGGACCTTG 1558

```
Db      564 AGGGAACCTTGGCGTCAGCCAGACAAAGAAAGACGAGGAGAAAAAGTCAGTTAG 505
      1559 CCGGCGGGAGAAAGACCGGGCCGAGAACTGATGATCGTCACTGTGTCCGACGACC 1618
      Db      504 CGACGAGTGTCAAGAGAGCTCGCCGAGAACTTAATGATTTGATCTTATCCGCGCGACC 445
      1619 TCAACAGCGCTCGCGGATCGGCTCGGTCCAGTGGCCCGGCTTTCGAGGTGGAACCT 1678
      Db      444 TCACACACATTTTCCATCAAAATCAATCAAGTCCCGAAACTACTTCCGTTGAAAGCT 385
      1679 AGCGCGCCCTGCACACAGCTGTGTGACATCCGCGGAGCGGTGCGGCCGCGACGACA 1738
      Db      384 AGAAGAACATGATCATGAGCTGTGACAAACATTCATCCCATCGGCCCAAGCTGGCG 325
      1739 CCGCGCGCTGCGTACCGCGCGCCCTTCCCGCGCGCTTCATGACGCGCGCGCCCAAGAGC 1798
      Db      324 GCGTCCAGGTTCTCGACGCGATGCTTCCACAGAGATCATGACGGGCGCTCCCAACTCA 265
      1799 GCACCATGAGATCATGACCGCGCTGAGAGAGAGGCCCC---CGGGCGCTTACTCCGGGG 1855
      Db      264 GAGCCCTGCAAAATCTCGACGCGCTGAGAGACACCGCAGCGCGCATATACTCCGACA 205
      1856 CGCTCGGATGTTTCCGCTCAGCGCGCGCGCGCAGCTCAGATCGTCAATCCGACCATCG 1915
      Db      204 GTCTGGGCTATGATATCGCGCAGCGGACCGTCGATCAGTCGGTGTATTCGTAACGATTG 145
      1916 TGTGCGCCGACGCGCAGCGGAGATTGCGCGTGGCGGGCGGATCGTGTCCCTCTCCGACC 1975
      Db      144 TAAAGTACGGAAGACGCTTGAGCTGAGTGTGAGAGGGCGGATTAATCGCTCAGCAGAG 85
      1976 AGGAGAGAGATTACACGAGACCGTGTAAAGGCCCGCGCATGATCACCGCCCTCGAGC 2035
      Db      84 CGAGAAAGATGAGATGAGTATGATGTGAAGCGATGTGTCTACGCGCTCGCTTA 25
      2036 GCAGCGCGCTGCGCGCGCGCCGAT 2059
      Db      24 GGGAAATCGGCGCAAGATGCTGAT 1

RESULT 14
COO30342/c      818 bp      mRNA      linear      EST 10-JUN-2004
LOCUS      ESR0808726 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
DEFINITION      kb Coccidioides posadasii cDNA clone C1FB368 3' end, mRNA sequence.
ACCESSION      COO30342
VERSION      COO30342.1 GI:48563318
KEYWORDS      EST.
SOURCE      Coccidioides posadasii
ORGANISM      Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE      1 (bases 1 to 818)
AUTHORS      Gardner,M.J. and Cole,G.T.
TITLE      Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL      Unpublished (2003)
COMMENT      Other_ESTs: ESR808727
            Contact: Gardner MJ
            The Institute for Genomic Research
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301 838 3519
            Fax: 301 838 0208
            Email: gardner@tigr.org.
FEATURES
            source
            1..818
            Location/Qualifiers
            /organism="Coccidioides posadasii"
            /mol_type="mRNA"
            /strain="C735"
            /db_xref="taxon:199306"
            /clone="C1FB368"
            /dev_stage="spherules"
            /lab_host="E. coli DH10B, T1 phage resistant"
```

```
ORIGIN
Query Match      8.0%; Score 164.8; DB 7; Length 818;
Best Local Similarity 55.1%; Pred. No.2.5e-18;
Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

      1439 CTTGCCGAGCGGTTCTCAGATCGGCGCGGACCGGCGGCTGAGTCCAAAGCCATCA 1498
      Db      624 CTTGCCGCGAGAGTTTATATCATGATGATGCTGATGCTGTGCTGTAATGAACCAATAA 565
      1499 AGGGAACCCGCCCGCGGCGGACCGCGAGAGAGACGAGTCCGCGCGACCTGG 1558
      Db      564 AGGGAACCTTGCCTCGCACCGACCAAGAAAGAAAGACGAGCGAGAAAAAGTCAGTTAG 505
      1559 CCGCGCGGAGAAAGACCGGGCCGAGAACTGATGATTCGACCTGTGTCGCAAGACC 1618
      Db      504 CGACGAGTGTCAAGAGCTCGCGGAGAACTTAATGATTTGATCTTATTCGCGCGACC 445
      1619 TCAACAGCGTTCGCGCATCGGCTCGGTCCAGTCCGCGCTTTCGAGGTGGAACCT 1678
      Db      444 TCACACACATTTCTCATCAAAATCAATCAAGTCCCGAAACTACTTCCAGTTGAAAGCT 385
      1679 AGCGCGCCCTGCACACAGCTGTGTGACATCCGCGGAGCGGTGCGGCCGCGACGACA 1738
      Db      384 AGAAGAACATGATCATGAGCTGTGACAAACCATTCATCCACATTCGCGCCCAAGTGGCG 325
      1739 CCGCGCGCTGCGTACGCGCGCCCTTCCCGGCGCTTCATGACCGGCGGCCCAAGAAC 1798
      Db      324 GCGTCCAGGTTCTCGACGCGATCTTCCACAGAGATCAATCCACATTCGCGCCCAAGTGGCG 265
      1799 GCACCATGAGATCATGACCGCGCTGAGAGAGAGGCCCC---CGGGCGCTTACTCCGGGG 1855
      Db      264 GAGCCCTGCAAAATCTCGACGCGCTGAGAGACACCGCAGCGCGCATATACTCCGACA 205
      1856 CGCTCGGATGTTTCCGCTCAGCGCGCGCGCGCAGCTCAGATCGTCAATCCGACCATCG 1915
      Db      204 GTCTGGGCTATGATATCGCGCAGCGGACCGTCGATCAGTCGGTGTATTCGTAACGATTG 145
      1916 TGTGCGCCGACGCGCAGCGGAGATTGCGCGTGGCGGGCGGATCGTGTCCCTCTCCGACC 1975
      Db      144 TAAAGTACGGAAGACGCTTGAGCTGAGTGTGAGAGGGCGGATTAATCGCTCAGCAGAG 85
      1976 AGGAGAGAGATTACACGAGACCGTGTAAAGGCCCGCGCATGATCACCGCCCTCGAGC 2035
      Db      84 CGAGAAAGATGAGATGAGTATGATGTGAAGCGATGTGTCTACGCGCTCGCTTA 25
      2036 GCAGCGCGCTGCGCGCGCGCCGAT 2059
      Db      24 GGGAAATCGGCGCAAGATGCTGAT 1

RESULT 15
COO34510/c      820 bp      mRNA      linear      EST 10-JUN-2004
LOCUS      ESR812894 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
DEFINITION      kb Coccidioides posadasii cDNA clone C1FB103 3' end, mRNA sequence.
ACCESSION      COO34510
VERSION      COO34510.1 GI:48571664
KEYWORDS      EST.
SOURCE      Coccidioides posadasii
ORGANISM      Coccidioides posadasii
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
            Onygenales; mitosporic Onygenales; Coccidioides.
REFERENCE      1 (bases 1 to 820)
AUTHORS      Gardner,M.J. and Cole,G.T.
TITLE      Analysis of gene expression in Coccidioides posadasii mycelia and
            spherules via expressed sequence tags
JOURNAL      Unpublished (2003)
```

COMMENT

Other_ESTS: EST812895
Contact: Gardner MJ
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@igr.org.

FEATURES

source

Location/Qualifiers
1..820
/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="C1F8U13"
/dev_stage="spherules"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_lib="Coccidioides posadasii spherule cDNA library,
0.5 to 5.3 Kb"
/note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
Kb "

ORIGIN

Query Match 8.0%; Score 164.8; DB 7; Length 820;
Best Local Similarity 55.1%; Pred. No. 2.5e-18;
Matches 344; Conservative 0; Mismatches 277; Indels 3; Gaps 1;

```
QY 1439 CCTCGCCCGAGCGGTTCTCAGCATCGGCGCGACGCGCGGCTTCGAGTCCAGCCCATCA 1498
DB 624 CGTCGCCGAGAGGTTATATTCATAGATCTGATGCGCTGTAATGAATCAATATA 565
QY 1499 AGGGGACCCGCCCCCGGGCGGACCGCGAGAGGAGACGAGCGGCTCCGCGCCGACCTGG 1558
DB 564 AGGGAACCTTGCGCTGTCAGCCAGACAAAGAGAAGACGAGCGGAGAAAAGTCACTTAG 505
QY 1559 CCGGCGGAGGAAGACCGGGCCGAGACCTGATGATCTGCACTGCTGTCGACGACC 1618
DB 504 CGACGATGTCAGAGGACTCGCCGAGAACTTAATGATTCGATCTTATCCGCGCGACCC 445
QY 1619 TCACAGCGCTCGCGGATCGGCTCGCTCCAGTCCGCGCGGCTTCGAGGTGAGACCT 1678
DB 444 TCACAAACATTTTCTCAATAAATCAATCAAGGTCCGAAATCACTCAAGTTGAAAGCT 385
QY 1679 ACCGCGCCGTGACACGAGTGTGACCATCCGGGAGCGGCTGCGGCGCGGACCGACA 1738
DB 384 ACGAACAGTACATCAAGTCTGTCAACACATTCATCCACATCGCCCCCAAGCTGGGG 325
QY 1739 CCGCGCGCTGCGTACGCGCGCTTCCCGCGGCTTCATGACCGCGCGGCCCAAGAGC 1798
DB 324 GCGTCGAGGTTTCGAGCGATGCTCCCAACAGATTCATGACGCGCGCTCCCAACTCA 265
QY 1799 GCACATGAGATCATGACCGGCTGAGGAGAGGCCCC--CGGGGCGTCTACTCGGGG 1855
DB 264 GAGCCGTCAAAATCTTGAGAGGCTTGAGAGAGCACCGGCGGCGGATATCTCTCGGCA 205
QY 1856 CGCTCGGATGTTGCGCTTCAGCGGCGCGCCGACCTCAGCATGTCATCCGACCATCG 1915
DB 204 GTCTGGGCTATGATGCGCAGCGGACCGTCGATCATGTCGTTATCCGATGATG 145
QY 1916 TGCTGCGCAGCGGCGAGCGGAGTTGCGGCTCGGCGGCGGATGCTGCTCTCCGACC 1975
DB 144 TAAAGTACGGAAGCAGCTTAGCTGAGGAGGCGGATTACTTGCTCAGCGAGG 85
QY 1976 AGGAGGAGGAGTTCACCGAGACCGGTGTAAGGCCCGCGCATGTCACCGCCCTGACG 2035
DB 84 CGAGAAAGATGAGATGAGTCAATGATGAGGCGAATCTGTTCTACGCGCGTGCCTA 25
QY 2036 GCAGCGCGCTGCGCGGCGCCGAT 2059
DB 24 GGGAAATCGGCGCCAGATGCTGAT 1
```

Job time : 7804.14 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 05:31:11 ; Search time 10157.2 Seconds
(without alignments)
3272.568 Million cell updates/sec

Title: US-10-089-514-2
Perfect score: 3573
Sequence: 1 MRTLIDNVDSTHNLFOYL.....VVRAMVTLADGSAAVAGAR 686

Scoring table: BLOSUM62
Xgapop 10.0 , Ygapext 0.5
Ygapop 10.0 , Fgapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-O=C922/USFTO.spool_n/US10089514/runat_04102005_105744_8033/app.query.fasta_1.1621
-DB=GenEmbl -QFMT=fastap -SUFFIX=cye -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORM=gxt -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514.@CCN 1.1.8782.@runat_04102005_105744_8033 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.*
1: gb_ha:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3573	100.0	2061	BD178313	BD178313 Transform
2	3573	100.0	2061	BD093914	BD093914 Transform
3	3573	100.0	5251	AB116234	AB116234 Streptomy
4	3127.5	87.5	2660	SVU21728	U21728 Streptomyce

5	3127.5	87.5	14159	1	AF262220	Streptomy
6	1816	50.8	2220	6	ARI98361	Sequence
7	1816	50.8	4740	1	SPU60417	Streptomyce
8	1766	49.4	4607	1	STMABA	
9	1766	49.4	39314	1	SGR300302	
10	1766	49.4	158203	1	AY310323	Streptomy
11	1635.5	44.8	110000	1	AP006618_17	
12	1606	44.9	333550	1	AP003593	Streptomy
13	1600.5	44.8	29132	1	STH575648	
14	1596.5	44.7	283300	1	AP005025	Streptomy
15	1541	43.1	300750	1	AP005217	Streptomy
16	1487	41.6	1860	6	BD163312	Novel pol
17	1487	41.6	1860	6	AX121195	Sequence
18	1487	41.6	1983	6	AX064177	Sequence
19	1487	41.6	1983	6	AX064419	Sequence
20	1487	41.6	2005	6	AX764213	Sequence
21	1487	41.6	333150	1	AP005277	Corynebact
22	1487	41.6	349459	1	BX927151	Corynebact
23	1487	41.6	349980	6	AX127146	Sequence
24	1475.5	41.3	349944	1	BX571871	Phototrab
25	1475.5	41.3	349980	6	AX770909	Sequence
26	1324	37.1	3174	8	AY425708	Lycopersi
27	1288	36.0	2760	6	AX765866	Sequence
28	1288	36.0	2760	8	BT002702	Arabidops
29	1288	36.0	3084	8	AY096797	Arabidops
30	1244	34.8	55839	1	AJ586576	Xanthomon
31	1228	34.4	26309	6	CQ363747	Sequence
32	1228	34.4	110000	1	AE017283_21	Continuation (22 o
33	1118	31.3	104045	2	AP006497_8	Continuation (9 of
34	1118	31.3	110000	2	AP006497_7	Continuation (8 of
35	1096.5	30.7	4496	6	A48326	Sequence 4
36	1096.5	30.7	4496	6	ARI98356	Sequence
37	1068.5	29.9	349895	1	BX248359	Corynebact
38	1029	28.8	38829	2	AC151926	Phaeodact
39	937.5	26.2	298050	1	AP003189	Clostridi
40	930	25.3	11639	1	AE010477	AE010477 Pseudobacte
41	904.5	25.3	107097	8	AC005727	Arabidops
42	897	25.1	110000	8	CR382127_05	Continuation (6 of
43	859.5	23.1	10226	1	AE006363	Continuation (13 o
44	837	24.4	110000	1	AE016822_12	Continuation (13 o
45	834	23.3	305050	1	AL596173	Listeria

ALIGNMENTS

RESULT 1
BD178313
LOCUS
DEFINITION
Transformation producing PR1022 substance and process for producing the same and novel biosynthetic gene.
ACCESSION
BD178313
VERSION
BD178313.1 GI:30015578
KEYWORDS
WO 02077244-A/1.
SOURCE
Streptomyces venezuelae
ORGANISM
Streptomyces venezuelae
REFERENCE
Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.
1 (bases 1 to 2061)
JOURNAL
Transformation producing PR1022 substance and process for producing the same and novel biosynthetic gene
Patent: WO 02077244-A 1 03-OCT-2002;
MEIJI SEIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI

COMMENT

OS Streptomyces venezuelae
PN WO 02077244-A/1
PD 03-OCT-2002 WO 2002JP002782
PF 22-MAR-2001 JP 01P 082227
PI KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI
PC C12N15/61, C12N1/15, C12P17/14, C12N1/15, C12R1/645, C12P17/14,

PC C12R1.645)
CC Transformant producing PF1022 substance and process for CC
CC producing the same
CC and novel biosynthetic gene
FH Key Location/Qualifiers
FT CDS (1)..(2058).

FEATURES
source 1..2061
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"

ORIGIN

Alignment Scores:

Pred. No.: 2..06e-166 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0

US-10-089-514-2 (1-686) x BD178313 (1-2061)

QY 1 MetATGThrLeuLeuLleAspSerPheThrHisAsnLeuPheGlnTyrIle 20
DB 1 ATGCGACGCTTCTGATCGACAACTACGACTCGTTACCCACCAACTGTTCCAGTACATC 60
QY 21 G1G1A1A1A1ThrG1Yg1InProProVal1Val1ProAsnAspAlaAspTrpSerArgLeu 40
DB 61 GGGAGAGCCACCGGGACACCCCGTGTGTGTCACCAACGACCCGACGTGTGCGGCTG 120
QY 41 ProValG1uAspPheAspAla1leVal1SerProG1YProG1YSerProAspArgGlu 60
DB 121 CCGGTGAGAGACTTCACGCGGATGTCGTGTCCCGGGCCCGGACGCCGACCGGGAA 180
QY 61 ArgAspPheG1Y1leSerArgArgAla1leThrAspSerG1YLeuProVal1leuG1YVal 80
DB 181 CGGAGCTTCGGAATCGACCGCGCGGAGATCACCGACAGCGGCTGCGCGCTCGCGCGTC 240
QY 81 CysLeuG1YHisGlnG1Y1leAlaG1nLeuPheG1Yg1YThrVal1G1YleuAlaProGlu 100
DB 241 TGCTTGAGCCACCGAGGACATCGCCAGCTTCTGGGAGAACCGTGGGCTCGCCCGGAA 300
QY 101 ProMetHisG1YArgVal1SerGluVal1ArgHis1eThrG1Yg1uAspVal1PheArgG1YLeu 120
DB 301 CCATGCAAGGCGGGGTCTCCGAGGTGCGGACACCGGCGAGGACGTCTTCCGGGCGCTC 360
QY 121 ProSerProPheThrAlaVal1ArgTYRHisSerLeuAla1aThrAspLeuProAspGlu 140
DB 361 CCTCGCCGCTTCAACCGCCGCTGCTACCACTCCCTGGCGGACCACTCCCGACGAG 420
QY 141 LeuGluProLeuAlaTrpSerAspAspG1YVal1MetG1YleuArgHisArgGluYrs 160
DB 421 CTCGAACCCCTCGCTGAGACGACGACGAGGAGTGTATGAGGCTTGCGGACCCCGAAG 480
QY 161 ProLeuTrpG1YVal1GlnPheHisProGluSer1eG1YSerAspPheG1YArgGluIle 180
DB 481 CCGCTGTGGGGCTTCCAGTTCCACCGGAGTCCATCGGACGACACTTCCGGCGGGAGATC 540
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisG1YAlaAsp 200
DB 541 ATGGCAAACTTCGCGACCTCGCCCTCGCCACCAACGGGACAGGCGGACCGGGCGAC 600
QY 201 SerProTYrGluLeuHisVal1ArgArgVal1AspVal1leuProAspAlaG1uG1YVal1Arg 220
DB 601 TCCCTGTAGAACTTCCACGTGCGCGCGCTGACGTCTGCGGACCCGAAAGGTAACG 660
QY 221 ArgG1YCysLeuProG1Yg1uG1YThrPheTrpLeuAspSerSerVal1leuGlu 240
DB 661 CCGGGCTGCTGCGCGGACGAGGACACAGTTCGTGGCTGACAGGACGTCCGTCCTCGAA 720
QY 241 G1YAlaSerArgPheSerPheLeuG1YAspAspArgG1YProLeuAlaG1uTYrLeuThr 260

DB 721 GGGCGCTCGCGCTTCTCTCTCGGCGACGACCGGCGCCGCTCGCGAGTACCTCAC 780
QY 261 TYrArgVal1AlaAspG1YVal1Val1SerVal1ArgG1YSerAspG1YThrThrArgTrp 280
DB 781 TACCGGTGGCGGACGGGCTGCTTCCTCGCGGCTCCGAGGACCAACGACCGGAG 840
QY 281 ArgArgProPhePheAsnTYrLeuGlnGlnGlnLeuGluArgArgVal1ProVal1Ala 300
DB 841 CCGCGCGCTTCTTCAACTACCTGGAGGAGCAGCTGAAACGCGGAGGCTCCCGTCGCG 900
QY 301 ProGluLeuProPheG1YleuAsnLeuG1YTYrVal1G1YTYrLeuG1YTYrGluLeuYrs 320
DB 901 CCGGACCTGCTTCCAGTTCAACTCGGCTAGTCGAGTACCTCGGCTTACGAGCTGAAG 960
QY 321 AlaGluThrThrG1YAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 961 GGGAGACCAACGGGACACCCCGGACCGGCTCCCGCACCCGACCCGCTTCTTCTTC 1020
QY 341 AlaAspArgAla1leAlaLeuAspHisGlnG1YCysCysTYrLeuLeuAlaLeuAsp 360
DB 1021 GCGACCGGCGCATGCGCTTCGACCAACGAGAAAGGTGCTGCTACCTGCTGGCCTCGAC 1080
QY 361 ArgArgG1YHisAspAspG1YAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr 380
DB 1081 CCGCGGGGACAGACGACGCGCGCGCTGCTGCGGAGACGCGCCGACACCTTCAC 1140
QY 381 G1YleuAlaVal1ArgAlaProAlaGluProThrProAlaMetVal1PheG1Y1leProGlu 400
DB 1141 GGGCTGGCTCGCGCCCGGCGGACCGGACCCCGCATGCTCTTGGAGATCCCCGAG 1200
QY 401 AlaAlaAlaG1YPheG1YProLeuAlaArgAlaArgHisAspAlaAspAlaTYrLeuYrs 420
DB 1201 GGGCGCGCGCTTGGCGCCCTCGGCGCGCGCGGCGGACGACGACGACGACCTACCTCAAG 1260
QY 421 Arg1leAspGluCysLeuYrsG1Y1leArgAsnG1Yg1uSerTYrG1Y1leCysLeuThr 440
DB 1261 CCGATCGAAGATGCTTCAAGAGATCCCGCAACGGGAGTCTGACGAGATCTCGCTGACC 1320
QY 441 AsnMetVal1ThrAlaProThrGluAlaThrAlaLeuProLeuTYrSerAlaLeuArgAla 460
DB 1321 AACATGTACCGCGCGGACCGAGCGAGCGGCGCTGCGCTTACTCCGCGCTCGCGCC 1380
QY 461 1leSerProVal1ProTYrG1YAlaLeuLeuGluPheProGluLeuSerVal1leuSerAla 480
DB 1381 ATCAGCCCGCTCCCGTACGCGCGCGCTCGATTCGCCGAATGTCGCTGACGCGCC 1440
QY 481 SerProGluArgPheLeuThr1leG1YAlaAspG1YG1YVal1GluSerTYrPro1leYrs 500
DB 1441 TCGCCCGAGCGGTTCTTCAAGATCGCGCGGACGCGCGGCTGAGTCCAAAGCCCATCAAG 1500
QY 501 G1YThrArgProArgG1Yg1YThrAlaG1uG1YAspGluArgLeuArgAlaAspLeuAla 520
DB 1501 GGGACCGCGCCCGGGGCGGACCGGAGAGAGAGAGAGCGGCTCGCGCCGACCTGGGCG 1560
QY 521 G1YArgGluYrsAspArgAlaGluAsnLeuMet1leVal1AspLeuVal1ArgAsnAspLeu 540
DB 1561 GCGCGGAGAAAGACCGGCGCGAAGACCTGATGATGTCGACTGGTCCGGAACGACCTC 1620
QY 541 AsnSerVal1CysAla1leG1YSerVal1HisVal1ProArgLeuPheGluVal1GluTYr 560
DB 1621 AACAGGTGTGGCGCATCGGCTCCGTCACGTCGCGCGCTTTCGAGGTGAGACCTTAC 1680
QY 561 AlaProVal1HisGluLeuVal1SerThr1leArgG1YArgLeuAspProG1YThrSerTrp 580
DB 1681 GGGCGCGTACACGCTGTGTGACCACTCGGAGAGGCTCGCGCGCCGACGACAC 1740
QY 581 AlaAlaCysVal1ArgAlaAlaPheProG1Yg1YSerMetThrG1YAlaProYrsYrsArg 600
DB 1741 GCGCGCTGTACGCGCGCGCTTCCCGCGGCTCATACCGGCGCGCCGCAAGAGCCG 1800
QY 601 ThrMetGlu1le1leAspArgLeuGlnG1YProArgG1YVal1YrsSerG1YAlaLeu 620
DB 1801 ACCATGAGATCATGACGCGCTGAGAGAGAGCCCGGGGCGTCTACTCGGGGCGCTC 1860

QY 621 G1YrPhhAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
DB 1861 GGAGGCTTCGGCCCTCAACGGGGCCGCCGACCTCGATCGATCCGACCATCTGCTG 1920
QY 641 AlaAspGlyAlaGlnAlaGlnPheGlyValGlyAlaIleValSerLeuSerAspGlnGlu 660
DB 1921 GCCGACGGCCAGCGGAGTTCCGCGTCGCGGGGCGATCGTCTCTCTCCGACCAAGAG 1980
QY 661 GlnGlnPheThrGlnThrValValValAlaArgAlaMetValThrAlaLeuAspGlySer 680
DB 1981 GAGAGTTTCACCGGACCGGTGTAAAGCCCGCCATGTCACCGCTTCGACGCGAGC 2040
QY 681 AlaValAlaGlyAlaArg 686
DB 2041 GCCGTGGCGGGCGCCCA 2058
RESULT 2
LOCUS BD093914
DEFINITION Transforms that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.
ACCESSION BD093914
VERSION BD093914.1 GI:22639502
KEYWORDS WO 0123542-A/1.
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
REFERENCE 1 (bases 1 to 2061)
AUTHORS Yana, K., Okakura, K., Yasuda, S., Watanabe, M., Miyamoto, K., Mido, N. and Murakami, T.
TITLE Transforms that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes
JOURNAL Patent: WO 0123542-A 1 05-Apr-2001;
MEIJU SEIKA KAISHA LTD, KOJI YANAI, KAZU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAOKI MIDO, TAKESHI MURAKAMI
COMMENT OS Streptomyces venezuelae
PN WO 0123542-A/1
PD 05-APR-2001
PR 29-SEP-2000 WO 2000JP06783
PI 29-SEP-1999 JP 99P 276314
PI KOJI YANAI, KAZU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO
PI NAOKI MIDO, TAKESHI MURAKAMI
PC C12N15/09, C12N5/10, C12P21/02, C07K11/00// (C12P21/02, C12R1:645)
CC
FH
FT CDS
FEATURES
source Location/Qualifiers
1..2061
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
ORIGIN
Alignment Scores:
Pred. No.: 2,066-166 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-089-514-2 (1-686) x BD093914 (1-2061)
QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
DB 1 ATGCGACACGCTTCTGATCGAACACTACGACTCGTTCAACCAACCTGTTCCAGTACATC 60
QY 21 GlyGlnAlaThrGlyGlnProProValValValProAsnAspAlaAspTyrSerArgLeu 40
DB 61 GCGAGGCGCACCGGGCAACCCCGCTGCTGTGCGCAACGACCGCGACTGTGTGCGGCTG 120

QY 41 ProValGlnAspPheAspAlaIleValIleSerProGlyProGlySerProAspArgGlu 60
DB 121 CCGGTGAGGACCTTCGACGGGATGCTGTCTCCCGGAGCCCGGACCGGAGAA 180
QY 61 ArgAspPheGlyIleSerArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
DB 181 CGGAGCTTCGGAATTCAGCCGCCCGGCGATACCGACGCGGCTGCCGCTCTCGGCTC 240
QY 81 CysLeuGlyYHIsglnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 241 TGCCCTGGGACACGAGGACATCGCCAGCTTCGGGCGAACCCTGGGCTCGCCCGGAA 300
QY 101 ProMetHisGlyValArgValSerGlnValArgHisThrGlyGlnAspValPheArgGlyLeu 120
DB 301 CCATGACGAGCGGGGCTTCGAGGTGGGACACCGGCGAGGACGTTCTTCGGGGCTTC 360
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140
DB 361 CCTCGCGCTTCACCGCGCTGCGCTACCACTCCCTGGCGGACGACCTCCCGAGAG 420
QY 141 LeuGlnProLeuAlaTyrSerAspAspGlyValIleMetGlyLeuArgHisArgGlyLys 160
DB 421 CTCGAACCCCTCGCTGAGCGACGACGAGGGGTGTGTCATGGGCTCGCGGACCGCGAGAG 480
QY 161 ProLeuThrGlyValGlnPheHisProGlnSerIleGlySerAspPheGlyValArgGluIle 180
DB 481 CCGCTGTGGGGCTTCGACGTTCCACCGGAGTTCATGGCGACCTTCGCGCGGAGATC 540
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyAlaAsp 200
DB 541 ATGGCCAACTTCGCGACCTCGCGCTTCGCCACACCGGCGACGCGGCGGCGGCGG 600
QY 201 SerProTyrGlyLeuHisValAlaArgArgValAspValLeuProAspAlaGlnGlyValArg 220
DB 601 TCCCGCTGACCACTCCACGCGCGCGCTGCACTGCTCGGAGCCGGAAGAGTACGC 660
QY 221 ArgGlyCysLeuProGlyGlyGlyThrThrPheThrPheLeuAspSerSerValLeuGlu 240
DB 661 CGCGGCTGCTGCGCGCGGCGGACGACGACGTTGTGCTGACGACGACCTCGCTCGAA 720
QY 241 GlyAlaSerArgPheSerPheLeuGlyValAspAspArgGlyProLeuAlaGlyTyrLeuThr 260
DB 721 GCGGCTCGCGCTTCCTCTCTCTCGGCGAGCGAGCGGCGGCTCGCGGAGTACTAC 780
QY 261 TyrArgValAlaAspGlyValIleSerValArgGlySerAspGlyThrThrThrArgThr 280
DB 781 TACCGGTGCGCGACGCGCGCTCGCTCGCGGCTCGGACGCGACGACCGGAGCG 840
QY 281 ArgArgProPhePheAsnTyrLeuGlnGlnGlnLeuGlnArgArgValProValAla 300
DB 841 CGCGCGCGCTTCCTCAACTACCTCGAGGAGCAGCTGGAACGCGGAGGCTCGCGCG 900
QY 301 ProGlnLeuProPheGlnPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGlyLeuLys 320
DB 901 CCGGAACCTGCTTCAGATTCACTCGGCTACGCTGAGCTACCTCGGCTACGAGCTGAG 960
QY 321 AlaGlnThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 961 GCGGAGACACACCGGCGACCCCGGCGACCGGCTCGCGACCGCGGCTTCCTTC 1020
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCysCysTyrLeuLeuAlaLeuAsp 360
DB 1021 GCGGACCGGCGCTTCCTCGACGACGACGAGGAGGCTGTGCTACCTGTGCGCTCGAC 1080
QY 361 ArgArgGlyYHIAspAspGlyAlaArgAlaTyrPheArgGlyThrAlaGlnThrLeuThr 380
DB 1081 GCGGAGGACACACGAGCGCGCGCTGCTGCGGAGAGCGGCGGACCTTCAC 1140
QY 381 GlyLeuAlaValArgAlaProAlaGlnProThrProAlaMetValPheGlyIleProGlu 400
DB 1141 GGCCTGCGGCTCGCGGCGCGGCGGAGCGGACCGCGGCAATGCTTCGGGATCCCGAG 1200

QY 401 AAlAlAAlAglyPhegylProleuAlaArgAlaAspIysAspAlaTyrLeuIys 420
| | | | |
Db 1201 GCGCGCGCGCTTGGCCCGCCCGCGCGCGCCACAGGACGCTACCTCAAG 1260
| | | | |
QY 421 ArgIAspGluCysLeuIysGluIleArgAsnGlyIleSerTyrGluIleCysLeuThr 440
| | | | |
Db 1261 CGCATGACGAGTGGCTTCAAGAGATCCGACAGCGGAGTGTGACGATCTCCGAGC 1320
| | | | |
QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
| | | | |
Db 1321 AACATGGTCAACCGCGCGACGAGGAGCGCCCTGCGCTTACTCCGCGCTGGCGGCC 1380
| | | | |
QY 461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
| | | | |
Db 1381 ATCAGCGCCCTCCGACGCGCGCTGTGACATCTCCCAACGTCGCTGAGCGGCC 1440
| | | | |
QY 481 SerProGluPhePheLeuThrIleGlyAlaAspGlyGlyValGluSerIysProIleIys 500
| | | | |
Db 1441 TCGCCGAGCGGTTCTTCAAGATCGCGCGCGAGCGGCGCTCAAGTTCAGGCCCATCAG 1500
| | | | |
QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
| | | | |
Db 1501 GGGACCGCGCCCGCGCGCGCGCACCGGAGAGAGAGAGAGCGGCTCGCGCGCGACCTGGCC 1560
| | | | |
QY 521 GlyArgGluIysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
| | | | |
Db 1561 GCGCGGAGAGAGAGAGCGCGCGCGGAACTGATGATGCTGACCTGCGCGCAAGCACTTC 1620
| | | | |
QY 541 AsnSerValCysAlaIleGlySerValIleValProArgLeuPheGluValGluThrTyr 560
| | | | |
Db 1621 AACAGGCTGCGCGATCGGCTCGCTCCGACGTGCGCCCGCTTTCAGAGTGGAGACTTAC 1680
| | | | |
QY 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
| | | | |
Db 1681 GCGCCCGGTGACACAGCTGTGTGACCATCCGGGAGAGCGTGGCGCGCGCACCAAGCAC 1740
| | | | |
QY 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPolysIysArg 600
| | | | |
Db 1741 GCGCGCTGTACGCGCGCGCTTCCCGCGCGCTCCATGACCGCGCGCGCAAGAGCGC 1800
| | | | |
QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
| | | | |
Db 1801 ACCATGGAGATCATCGACCGCTGAGAGAGAGCGCCCGGGCGGTCTACTCCGGGCGCTC 1860
| | | | |
QY 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
| | | | |
Db 1861 GATGTGTGCTCCCTCAGCGCGCGCGCGCGACTCAGCATCTCTCCGACCATCTGTCTG 1920
| | | | |
QY 641 AlaAspGlyGluAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGluGlu 660
| | | | |
Db 1921 GCGGACGCGCGAGCGCGGTTCGCGGTGCGCGCGCGCGCATCTGTCTCCGACGAGAG 1980
| | | | |
QY 661 GluGluPheThrGluThrValValIysAlaArgAlaMetValThrAlaLeuAspGlySer 680
| | | | |
Db 1981 GAGGAGTTCACCGAGACCGTGTGAAGGCCCGCGCATGTGTACCGCGCTCGAGCGGAGC 2040
| | | | |
QY 681 AlaValAlaGlyAlaArg 686
| | | | |
Db 2041 GCCGTGGCGGCGCGCGCA 2058
| | | | |
RESULT 3
AB116234 5251 bp DNA linear BCT 03-JUL-2004
LOCUS AB116234
DEFINITION Streptomyces venezuelae papa, papC, papB, ORFV genes for
4-amino-4-deoxychorismate synthase, 4-amino-4-deoxyprephenate
dehydrogenase, 4-amino-4-deoxychorismate mutase, hypothetical
protein, complete cds.
ACCESSION AB116234
VERSION AB116234
KEYWORDS AB116234.1 GI:47846865
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

REFERENCE
1
AUTHORS Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and Murakami,T.
TITLE Para-position derivatives of fungal anthelmintic cyclopeptides engineered with Streptomyces venezuelae antibiotic biosynthetic genes
JOURNAL Nat. Biotechnol. 22 (7), 848-855 (2004)
PUBMED 15184904
REFERENCE 2 (bases 1 to 5251)
AUTHORS Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and Murakami,T.
TITLE Direct Submission
SUBMITTED (01-AUG-2003) Koji Yanai, Meiji Seika Kaisha, Ltd., Microbiological Resources and Technology Laboratories, 788 Kayama, Odawara-shi, Kanagawa 2500852, Japan
(E-mail:koji.yanai@meiji.co.jp, Tel:81-465-37-5106, Fax:81-465-37-6397)
FEATURES
source
1..5251
Location/Qualifiers
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
651..2711
/gene="papa"
651..2711
/gene="papa"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="4-amino-4-deoxychorismate synthase"
/protein_id="BAD21140.1"
/db_xref="GI:47846866"
/translation="MRTLLIDNYDSFTNHLFOYIGATGQPVVVVNDMSRLPYED
PRAIVSRGSGPDRRDRGTSRAITTSGLPVLGVCCHGDIADLFGTGTGLAPRM
HRRVSEVRHTGSDVFRGLSPPTAYRYSLATTDLPDELEFLAKSDGVYMGRLRREK
PLMGVQHPHESIGSDPGRITMANFRLDALAHARHAGDSVELEHVRVDVLPDSE
VRRGCLPGEITFMIDSSVLEASRFSFLGDDRLAEYLTYRVAADVVSVRGSDGT
TTRTRRPFFENYLEOELERRRVPARELPPEFNLGYVGLYELKAKETTDPAHRSPP
DAFLPADRALIDHOGCCYLLALDRRGDDGARAMLETAETLGLAVRAPAEPT
AMVFGPEAAAGPGLAARHDKDAVYKRDICKEIRNGSEYETGLTMVAPAEPT
ALPLYSALRAISPPVYGALLEPPELSVLSASPERLTTGADGVSKPTKGRPRGT
AAEDERLRADLGRERDRAENIMTIDVLKNDLNSVCAIGSVAPRLFEVETAPVQL
VSTIRGRLPSTACVRAAPFGSGMTGAPKRTMEIIDRLEBGRGVSGALGWFA
LSGADLSIVRTIVLADQAEFGVGAIVLSIDQEEFTETVVKARAVMTALDGSAY
AGAR"
2708..3676
/gene="papC"
2708..3676
/gene="papC"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="4-amino-4-deoxyprephenate dehydrogenase"
/protein_id="BAD21141.1"
/db_xref="GI:47846867"
/translation="MSGFPRSYVVGSGAGVGMPAGLIREASRTLVVDLPVPPGRPD
ACTGVPTAPBPELAALRDADLVILVAHEVDLARAIVAPVTRMLMGALLADLSVRT
GMAALELAHAPGVQVGLNPMFAPAGTGRVVAIVVRDGRVYALRLIYVCGGGRP
VRLTAEHDRTAAQTALTHAVILSFGLLARLGVDRVALATAPRPHVLLALRLAR
LGSSPEVYGDIGRNPRAASARALAEALRSFALVGDPPDADAPGRADAPGHGPGC
DGAGNDGVFGELRLMLPBLAAGDHCQELFRTLHRTDDEGEKDR"
3673..3984
/gene="papB"
3673..3984
/gene="papB"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="4-amino-4-deoxychorismate mutase"
/protein_id="BAD21142.1"
/db_xref="GI:47846868"
/translation="MTEQNELQRLRAELDALDGLTLDVTRRRIDLGVRITARYKSRHV

PMNQGRVSLVKDRAARYAADHGLDESLVNLVDVITTEMCRVEDLVMSRESLTAEBR
R"
4065..5243
/gene="ORFV"
4065..5243
/gene="ORFV"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="BAD2143.1"
/db_xref="GI:47846869"
/translation="MATTTTAAITGRTAAQPPYMPDPAVDVTERELAGLPVLTPEP
EVDLNMGMALVVRGAALLVGGPCARFPEAVPDVRRVNDLQGLAATMRAGSGLP
AVALRGAQGVKPRSPFETESGARKKMPSGCAVNVPEPEAPARTPPRRLTLA
YTCSSQIVLDIRSMSGRPVLERYTTSHELLLPYEPVLRBAGHGTSGSAIFGMIG
ERTREDEGHAVLAQAVNPVGVKGTPTSDEDAVALSRLENGVGVGLTFTVRGA
KEVDELPLPVRAVARHGAIVMLCDPMHNGKLKAGHKTRLIEPMRAETVAEVRTIR
EHQWPAGLHLELTPDPTVECSBLDRPPRFTDYRSTCDRLNPEQASDWMVTHFLSL
"

ORIGIN

Alignment Scores:
Pred. No.: 5,72e-166 Length: 5251
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
US-10-089-514-2 (1-686) x AB116234 (1-5251)

QY 1 MetArgThrIleuLeuIleAspAsnTyrAspSerPheThrHisAsnIleuPheGlnTyrTle 20
DB 651 ATCGGACCGCTTCTGATCGACCACTAGACTCGTTCCACCAACCTGTTCCAGTCAATC 710
QY 21 G1yG1ua1aThrG1yG1nPProva1aIvalProbaAsp1aAAspTpsSerArgLeu 40
DB 711 GGGAGGCGCACCGGGCAACCCCGCTGCTGTCGCCAAGCCGCACTGTCGCGGCTG 770
QY 41 ProvalG1uaAspPheAspAlaIleValIleValSerProG1yProG1ySerProAspArg1u 60
DB 771 CCGCTCGAGGACTTCAGACGCGATCGTCGTCCCGGGCCCGGCAAGCCCGCAACCGGAA 830
QY 61 ArgAspPheG1y1IleSerArgArgAlaIleThrAspSerG1yLeuProvalIleuG1yVal 80
DB 831 CGGACTTCGGAATCAGCCCGCGGCGATCAACGACGCGCTGCGCTTCGCGGCTC 890
QY 81 CylLeuG1yHisG1ng1y1Ile1aG1nLeuPheG1yG1yThrValG1yLeuAlaProG1u 100
DB 891 TGCCTCGGCCACAGGGGATCGCCAGCTTTCGGGGAAACCGTCCGCTCGCCCGGAA 950
QY 101 Prometh1aG1yArgValSerG1uValArgHis1ThrG1yG1uAspVal1PheArgG1yLeu 120
DB 951 CCATGCAAGCGCGGGCTCCAGGTGCGGCAACCGGGGAGGACGCTTCGCGGGCTC 1010
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspG1u 140
DB 1011 CCTTCGCGCTTCAACCGCGCTGCTCACTCCCTGCGCGCACCACTCCCGGACGAG 1070
QY 141 LeuG1uProLeuAlaIlePheAspAspG1yValIleMerG1yLeuArgHisAspG1yVal 160
DB 1071 CTCGAACCCCTCGCTCGACCGACGAGGGGTGCTATGGGCTCGGCGACCGCGAAG 1130
QY 161 ProLeuTpsG1yValG1nPheHisProG1uSer1aG1ySerAspPheG1yArgG1u1e 180
DB 1131 CCGCTCGGGGCGCTCCAGTTCACCCGAGTTCATCGGACGACGACTTCGCGCGGAGATC 1190
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHis1aArgAlaArgArgHisG1yValaAsp 200
DB 1191 ATGGCAAACTTCGCGACCTCGCTCGCCACACCGGCGACGGGCGCACCGGGGCGGAC 1250
QY 201 SerProTyrG1uLeuHisValaArgArgValAspValLeuProAspAlaG1uG1uValArg 220

DB 1251 TCCCGGTAGCAACTCCACGTCGCCCGCGTCGAGCTGTCGCCGACCCGGAAGAGTACGC 1310
QY 221 ArgG1yCylLeuProG1yG1uG1yThrPheTpsLeuAspSerSerValIleuG1u 240
DB 1311 CGCGGCTGCTGCGCCCGGCGAGGACCAACGTTCTGCTGGAACAGACGCTCCGTCTCGAA 1370
QY 241 G1yAlaSerArgPheSerPheLeuG1yAspAspArgG1yProleuAlaG1uTyrIleuThr 260
DB 1371 GGGCGCTCGGCTTCTCTCTCGCGACGACCGGGCGCGCTCCCGGTACCTCAC 1430
QY 261 TyrArgValAlaAspG1yValIleSerValArgG1ySerAspG1yThrThrArgThr 280
DB 1431 TACCGGTCCCGCACCGGCTCGCTCCGTCGCGGCTCCGACGACACGACCCGAGC 1490
QY 281 ArgArgProPhePheAsnTyrIleuG1uG1uG1uLeuG1uArgArgValProvalAla 300
DB 1491 CGCGCGCTTCTTCAACTACCTGAGAGACGACTGAAACGCGGACCGGCTCCGCTGCGC 1550
QY 301 ProG1uLeuProPheG1uPheAsnLeuG1yTyrValG1yTyrIleuG1yTyrG1uLeuVal 320
DB 1551 CCGAACTGCTTTCAGATTCAACTCGGCTAGCTGCTACCTCGGCTACGAGCTGAG 1610
QY 321 AlaG1uThrThrG1yAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 1611 GCGGACACACCGGCGACCCCGGCGACCGGTCGCCGACCCGACCGCGCTTCTTTC 1670
QY 341 AlaAspArgAlaIleAlaLeuAspHisG1ng1yG1yCysArgTyrIleuLeuAlaLeuAsp 360
DB 1671 GCGGACCGGCGCATCGCTCGACACACGAAAGCTGCTGCTACCTGCTGCGCTCGAC 1730
QY 361 ArgArgG1yHisAspAspG1yAlaArgAlaIlePheLeuArgG1uThrAlaG1uThrLeuThr 380
DB 1731 CGCGGGGCGACGACACGCGCGCTGCTGCTGCGGAGACGCGCACGACCTCAC 1790
QY 381 G1yLeuAlaValArgAlaProAlaG1uProThrProAlaMerValPheG1yIleProG1u 400
DB 1791 GCGCTGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1850
QY 401 AlaAlaIleG1yPheG1yProleuAlaArgAlaIleArgHisAspIleAspAlaTyrIleuVal 420
DB 1851 GCGCGCGCGCGCTTGGCGCCCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1910
QY 421 ArgIleAspG1uCylLeuLeuG1uIleArgAsnG1yG1uSerTyrG1uIleCylLeuThr 440
DB 1911 CGATTCAGAGAGTCTCAAGAGATCCGCAACGCGGATCGATCCAGATCTGCTGACC 1970
QY 441 AsnMerValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
DB 1971 AACATGATCACCGCGCGACGAGGCGAGCGCGCTGCGCTCACTCCGCGCTGCGCGC 2030
QY 461 IleSerProvalProTyrG1yAlaLeuLeuG1uPheProG1uLeuSerValIleuSerAla 480
DB 2031 ATCAGCCCGCTCCGTAACGCGCCCTGCTCGAGTTCGCCGAATGTCGCTGAGCGCC 2090
QY 481 SerProG1uArgPheIleuThrIleG1yAlaAspG1yG1yG1uSerIleuProIleuVal 500
DB 2091 TCGCCCGAGCGGTTCTCTCAAGATCGCGCGCGAGCGGCGGTGAGTCCAAAGCCAAACAG 2150
QY 501 G1yThrArgProArgG1yG1yThrAlaG1uG1uAspG1uArgLeuArgAlaAspLeuAla 520
DB 2151 GGAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2210
QY 521 G1yArgG1uValAspArgAlaG1uAsnLeuMer1IleValAspLeuValArgAsnAspLeu 540
DB 2211 GCGCGGAGAAAGAACCGGCGAGAACCTGATGATCTGTGACTGTCGCGACACACCTC 2270
QY 541 AsnSerValCysAlaIleG1ySerValHisValProArgLeuPheG1uValG1uThrTyr 560
DB 2271 AACAGGTCTGCGCATCGGCTCCGTCCACGTCGCCGCTTCTGAGGTGAGACCTAC 2330
QY 561 AlaProvalHisG1uLeuValSerThrIleArgG1yArgLeuArgProG1yThrSerThr 580

Db 2331 GCCCCGTCACACGACTGCTGCACCACTCCGGGAGCGCTGGCGCCCGGACACGACACC 2390
Qy AlaAlaCyseValaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProTylsYsArg 600
Db 2391 GCGGCTGCGTACGCGCCGCGCTTCCCCGGGGGCTCCATGACCGGCGCCCAAGAACCGC 2450
Qy 601 ThMetGluIleIleAparGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
Db 2451 ACCATGAGATCATCAACCGCTCGAGAGAGCGCCCGGGGCGTCACTCCGGGCGCTC 2510
Qy 621 GlyTrpPheIleAseuSerGlyAlaAlaAseuSerIleValIleAargTrpIleValIleu 640
Db 2511 GGATGGTTCGCTTCAGCGCGCGCGGACCTGACATCGTCATCCGACCATCGTGTG 2570
Qy AlaAsePGlyAlaAlaGluPheGlyAlaGlyGlyAlaIleValSerLeuSerAsePGluGly 660
Db 2571 GCCGACGCGCACGCGGAGTTCGGCGTCCGGGGGCGCATCGTCTCCCTCCGACCGAGAG 2630
Qy 661 GluGluPheThrGlyIleValIleValAlaArgAlaMetValThrAlaLeuAsePGlySer 680
Db 2631 GAGGAGTTCACGAGACCGGTAAAGCGCGCGCATGTGTACCGCGCTCGACGCGCAGC 2690
Qy 681 AlaValAlaGlyAlaArg 686
Db 2691 GCCGTGCGGCGCGCCGA 2708

RESULT 4

SVU21728 2660 bp DNA linear BCT 22-NOV-1996
LOCUS Streptomyces venezuelae p-aminobenzoic acid synthase (pabAB) gene,
DEFINITION complete cds.
ACCESSION U21728
VERSION U21728.1 GI:1680048
KEYWORDS
SOURCE
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 2660)
AUTHORS Brown,M.P., Aidoo,K.A. and Vining,L.C.
TITLE A role for pabAB, a p-aminobenzoate synthase gene of Streptomyces
JOURNAL venezuelae ISB5230, in chloramphenicol biosynthesis
MEDLINE Microbiology 142 (Pt 6), 1345-1355 (1996)
PUBMED 96262706
REFERENCE 2 (bases 1 to 2660)
AUTHORS Brown,M.P.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1995) Biology, Dalhousie University, Halifax,
Nova Scotia B3H 4H6, Canada
COMMENT On Nov 21, 1996 this sequence version replaced gi:1680040.
FEATURES
source location/Qualifiers
1..2660
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ISP5230"
/db_xref="taxon:54571"
<1..507
/note="ORF1"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAB30312.1"
/db_xref="GI:1680048"
/translation="GSAEYVALPFLGHEHDLIRSKTGMIRKGSVLPADSTNIS
PMYITVAEYIGVDVTFIGMESTIGAAAMTIGPLTGEPLDRTDTSRLNGSPFOA
HREIVLLEPDEYVYVAMGLEPMWGVMAVDYDSHPAIVYSDLLVAHVODKGTARL
HURRLRL"
551..2660
/gene="pabAB"
651..2660
/gene="pabAB"
/codon_start=1

/transl_table=1
/product="p-aminobenzoic acid synthase"
/protein_id="AAB30312.1"
/db_xref="GI:1680048"
/translation="WRITLLINDYDSFTONTLFOYIGETATQCPVPPVNDADMSRLPLEDF
DAIVSPGSGSPDERDFEISRRATIDSGPLVGLGHLQGIQAQISAEPMHRSVEBR
HTGEIVFRGLPSPFTAVRHYSLAATDLDPLESEPLMSDGVVGLRHREKPLMGVQFP
PESIGDFRGREIMANFRDLALAHRRARDAADMGYELHYRVDVLPDAEVRRAACPA
EGATPELDSSSVLEGASPPSPFLGDDRGPLAEYLYTVADGVVSRGSDGTTTRDAATL
PSTLEOLEPSPAGPVAPDLPEFENIGYGVGLGELKERTGDPAPVAPHPDAAPLPAD
RAIALDHQGCCTLLADBRGHDDARAMLRETTATLTGLAARVPRPRAVRCVPE
AAAGPPLARAHDDASALRNSESIECTIMVTAFTBATLPYLSARRISPVPSG
ALLPEPLSVLSASPERFLTIGADGVSKPDKGRPRGAPAEEDERLRADLAGKED
RAENLMIVLVANDLNSVCAIGSVHVPPLFVEGDLAPYHQLVSTIRGLRPGTSTAAC
VAAAPFGSGMTCAPKRRPWEIIDRLKEEGRGVLPALGFWALSGAADLSIVRTIYLA
DGRARFGVGAIVLSLDGEERFQIVVARAAVVTALDGSAAVAGAR"

ORIGIN

Alignment Scores:

	Pred. No.:	Length:	Matches:
Score:	1,8e-144	2660	3127.50
Percent Similarity:	91.40%		618
Best Local Similarity:	90.09%		42
Query Match:	87.53%		17
DB:	1	Gaps:	3

US-10-089-514-2 (1-686) x SVU21728 (1-2660)

Qy 1 MetAArgThLeuLeuIleAsePantYrAsePheThrHiAseLeuPheGIntYrIle 20
Db 651 ATGCGACCGCTTCTGATCGACAACCACTGCTGTTCACCCAGAACCTGTTCAGATACATC 710
Qy 21 GlyIleAlaThGlyGlnProProValValIleProAsnAspAlaAspTrpSerArgLeu 40
Db 711 GCGAGGCCACCGGGGAGCCCTCCCTGCGCCACAGACGCGGACGTGTGCGGGCTG 767
Qy 41 ProValGluAsePheAspAlaIleValAseProGlyProGlySerProAparGlu 60
Db 768 CCCCTCGAGACTTCGACGCGATCGTGTCTCCCGGGGCGCGGACGCCGACCGAGNA 827
Qy 61 ArgAsePheGlyIleSerArgArgAlaIleThrAsePseGlyLeuProValLeuGlyVal 80
Db 828 CCGGACTTGAGATACGCGCGGCGGATACCGACGAGCGGCTCGCTCCGCGCTC 887
Qy 81 CyLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
Db 888 TGCCTCGGCGCACGAGGCGATCGCCAG-----CTTCGGCGGNA 926
Qy 101 PromethIleGlyArgValSerGlyValArgHiIleArgGlyAsePValPheArgGlyLeu 120
Db 927 CCATGCAAGCGCGGCTCCGAGGTCGGGACACCGGCGAGAGACGTTCCGGGGCTC 986
Qy 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAseLeuProAsePglu 140
Db 987 CCTCGCGCTTACCGCGCGCTGCGTCAACCTCCGCGCGCACCGACCTCCCGACGAG 1046
Qy 141 LeuGluProLeuAlaTrpSerAsePsePgluValIleMetGlyLeuArgHisArgGluYs 160
Db 1047 CTGAAACCCCTCGCTCGAGCGACGAGCGCGTGCATAGGCTGCGGACCGCGAGAG 1106
Qy 161 ProLeuTrpGlyValGlnPheHisPseProGluSerIleGlySerAsePseGlyArgGluIle 180
Db 1107 CCGCGATGGGCGTCCAGTTCACCGGAGTCCATCGGAGGAGACTTGGCGGGGAGATC 1166
Qy 181 MetAlaAsnPheArgAsePseValLeuAlaHisHisIleArgAlaArgArgHisGlyAlaAseP 200
Db 1167 ATGGCCAACTTCGCGACCTCCCTCGCGCACACCGGCGCAGTCCGCGCACCGCGCGAC 1226
Qy 201 SerProTyrGluLeuHisValArgArgValAsePValLeuProAsePAlaGluGluValArg 220
Db 1227 TGGGCTTACGAACCTCCAGCTGGCGCGCTCGACGTCGCTCCGAGCGCCAGAGGATACCC 1286
Qy 221 ArgGlyCyseLeuProGlyGluGlyThrThrPheTrpLeuAsePseSerValLeuGlu 240

Db	1287	CGCGCTGCCTGCGCGCGAGGCGCCACGTTCTTGCTGACGACGACTCCGTCTTCCAA	1346
QY	241	GIYVIAASERAPGPhSerPheLeuGIYAspAspArgGlyProLeuAlaGlyIYLeuThr	260
Db	1347	GCGCGCTGCGCGTCTCTCTCTCTCGCGACGACCGCGGCGCGCTCGCGAGTACTCAC	1406
QY	261	TYRArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr	280
Db	1407	TACCGCGTGCGCGAGGCGGTCTCTCTCGCTCGCGGCTCCGACGGGACCAACCGCGGAC	1466
QY	281	ArgArgProPhePheAsnTYrLeuGIYLeuGIYLeuGluArgArgValProValAla	300
Db	1467	GCGGAGACCTCTTACGCTACTGTAGAGACGCTCAACCGCGCGCGGTCTCGTCC	1528
QY	301	ProGluLeuProPheGluPheAsnLeuGIYTYrValGIYTYrLeuGIYTYrGlyIYLeuLys	320
Db	1527	CCCGACCTGCGCTTGAAGTTCACCTCGGACTACGTGCGCTACCTCGCTTACGAGTCAAG	1586
QY	321	AlaGlyThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe	340
Db	1587	GCGGAGACACACCGGCGACCCCGCACTACCGGCGCGCACCCCGACCGCGCTCTCTTC	1646
QY	341	AlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCysCysTYrLeuLeuAlaLeuAsp	360
Db	1647	GCGACCGCGCATTCGCCCTGCACCAACGAGAGGCTGCTCTACCTGCTG3CCTTCAC	1706
QY	361	ArgArgGlyHisAspAspGlyValaArgAlaTYrLeuArgGlyIYThrAlaGlyIYThrLeuThr	380
Db	1707	CGCGGCGGCGACGACGCGGCGCGCGCTGCGTGGGAGAGCGCGGACCACTTCAC	1766
QY	381	GlyLeuAlaValaArgAlaProAlaGluProThrProAlaMetValPheGlyIYLeuProGlu	400
Db	1767	GCGCTGCGCGTTCGCGTCCGCGCGCGACCGCGACCCCGCGACAGTCTTTCGCGGTCCCGAG	1826
QY	401	AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTYrLeuLys	420
Db	1827	GCGGCGCGCGCTTCGCGCGCGCTGCGTGCAGCGCACGACGACGAGAGCGCC-----	1877
QY	421	ArgIleAspGlyCysLeuLysGlyIleArgAsnGlyIYLeuSerTYrGlyIYLeuCysLeuThr	440
Db	1878	-----TCGGCGCTCCGCAACCGCGAGCTGTACGAGATTCGCTGCTAC	1919
QY	441	AsnMetValThrAlaProThrGlyAlaThrAlaLeuProLeuTYrSerAlaLeuArgAla	460
Db	1920	AACATGTGTACCGCGCGACGAGCGAGCGCGCTGCTCACTCGCGCTGCGCGC	1979
QY	461	IleSerProValProTYrGlyValaLeuLeuGluPheProGluLeuSerValLeuSerAla	480
Db	1980	ATCAGCCCCGCTCCGCTCGGCGCGCTGTCTCAGTTCGCCGAGCTGCGTCTCAGCGCC	2039
QY	481	SerProGluArgPheLeuThrIleGlyAlaAspGlyValaGlyIYLeuSerValProIleLys	500
Db	2040	TCGCGCGAGCGCTTCTCAAGATCGGCGCGCACCGCGCGCTCGAGTCCAMCCCATCAG	2099
QY	501	GlyIYThrArgProArgIYGIYTYrThrAlaGluGluAspGlyArgLeuArgAlaAspLeuAla	520
Db	2100	GGGACCGCGCCCGCGGGGCGCACCGCGGAGGAGCGAGCGGCTCCGCGCGGACCTGGCC	2159
QY	521	GlyIYrGluGlyAspArgAlaGlyIYAsnLeuMetIleValaAspLeuValaArgAsnAspLeu	540
Db	2160	GCGCGGAGGAGGACCGGCGCGAACAACCTGTGATCGTCGACCTGTGTCGCGACGACTC	2219
QY	541	AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlyIYValaGlyIYThrTYr	560
Db	2220	AACAGCGCTCTCGGATCGGCTCCGTCCACGTGCGCGCGCTCTTTCAGAGTGGAGACTTC	2279
QY	561	AlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyIYThrSerThr	580
Db	2280	GCGCGCGGACCAAGCTGTGTCCACATCCGGGAGCGGCTGCGGCGCGGACCAAGCAC	2339
QY	581	AlaIaCysValaArgAlaAlaPheProGlyGlyIYSerMetThrGlyIYAlaProLysValArg	600

Df	2340	GCCGCGTCGGCAGCGGCCGCTTCCCGGGGCGGTCTCATACCGGCGGCCCAAGAAGCA	2398
Oy	601	Thnrtwglutllellaesparagluengluuglyproargilyaltrysergilyalaenu	620
Df	2400	CCCATGGAGATCATCGACCGCTCGAAGAGAGCCCCCGGGCGGTCTTAACCGGGGCGCTC	2455
Oy	621	GlYtTPheAlaleusergilyalAlaAspieuserilleallleaqThrillealleu	640
Df	2460	GGATGGTTTGCCCTCACCGGGCGGCCGACTCGACATCGTCATCCGCACATCGTCTG	2519
Oy	641	AlaaspglygcinalagluPheglyValGlyglylalleValseaserAspglyglu	660
Df	2520	GCCACGCGCGGGCGCGCTTCGGCGCGGGCGAGTCGTCTCTTCCGACCAGGAG	2579
Oy	661	GlugluPheHrgrjurnrrvallylaAlaArgAlametValThrAlaleuasPglyser	680
Df	2580	GAGAGATTACGACGACCGGTGTAAAGCCCGGCCCATGTGTCACCGCTTCAGCGGACG	2639
Oy	681	AlaValAlaaglalyalaArg	686
Df	2640	GCAGTGGCGGGCGGCCGA	2657
RESULT 5			
AF262220			
LOCUS	AF262220	14159 bp	DNA linear BCT 09-JAN-2002
DEFINITION	Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,		
ACCESSION	AF262220		
VERSION	AF262220.2	GI:14290414	
KEYWORDS			
SOURCE	Streptomyces venezuelae		
ORGANISM	Streptomyces venezuelae		
REFERENCE	Bacteria; Actinobacterlia; Actinobacteridae; Actinomycetales;		
AUTHORS	Streptomycinase; Streptomycetaceae; Streptomyces.		
TITLE	1 (bases 1 to 14159)		
	He,J., Magarvey,N., Pirrae,M. and Vining,L.C.		
JOURNAL	The gene cluster for chloramphenicol biosynthesis in Streptomyces		
MEDLINE	venezuelae ISP5230 includes novel shikimate pathway homologues and		
PUBMED	a monomolecular non-ribosomal peptide synthetase gene		
	Microbiology 147 (Pt 10), 2817-2829 (2001)		
	21461106		
	11577160		
REFERENCE	2 (bases 1 to 14159)		
AUTHORS	He,J., Magarvey,N.A. and Vining,L.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford		
	St., Halifax, NS B3H 4U1, Canada		
REFERENCE	3 (bases 1 to 14159)		
AUTHORS	He,J., Magarvey,N.A. and Vining,L.C.		
TITLE	Direct Submission		
JOURNAL	Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford		
	St., Halifax, NS B3H 4U1, Canada		
REMARK	Sequence update by submitter		
COMMENT	On Jun 5, 2001 this sequence version replaced gi:10716943.		
FEATURES	Location/Qualifiers		
source	1..14159		
	/organism="Streptomyces venezuelae"		
	/mol_type="genomic DNA"		
	/strain="ISP5230"		
	/db_xref="taxon:54571"		
	/map="adjacent to pabAB"		
	233..3511		
	/function="essential for chloramphenicol biosynthesis"		
	/codon_start=1		
	/transl_table=11		
	/product="CmIG"		
	/protein_id="MAG21974.1"		
	/db_xref="GI:10716944"		
	/translation="MTAGSGPPAGRTRLPRLAVRDHSFLADGMDCVPNCTAPV		
	GTEPERVDPAAMRREFVARICGTALADLVAERDVTYTLDPRGVGCPAEALAD		
	LVTAYNSRPRLIDEVRLAYMRHTDGETIVFKESHLVLDLADVDLRQIRAHLAG		
	DTPPEGISGRHRHARTVEOYTLPADPAEVFRALGELPVPERKGIPTISASVEWL		
	PLNEGTIFELLSAVTGLVPLALGGGLVIQYPVSRMEFAIKGGYYEIKPLVVRPGLP		
CDS			

CDS
3498..66404
/note="non-ribosomal peptide synthetase: contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
/product="CmIH"
/protein_id="FAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITLFRHRKRLTAKRTWTYEELDRYSALRIID
AECPPGRRLVAGEHTAEAVVWALAAARSLAHTPMNPGLPAPRFEPAVADALV
CFERBALVRGEKAGLRALYAGDVWPTDPADAPADGTADBPASRAVYSIFTSSTGD
PKLVDSGHGKLMCRSLRLDITPDQDLPASLPSDVSSEIIGTYAGATLVVP
VRDQASWIGSVSRHIAHGCDLMLSPSYARLDPEARSRIRVEFGSGLATGEXDK
AARYSRVNAVPTTEATVCSLAEITSYPTSTCTPYDGPATVYRDDSGDHATGCE
LVIVGQVALGVAGSSPAENEVFYVDGSAVATGVNLSISDDELTVLRIDEQIKR
LGRHVLAAHGSTLSRHLGREVALVRQDAILTVADGATBESLMARIDLVPWE
APRLVLVDLPLTSGKVDRSALREILASPAAGPHGTGDEDAELRRLDVTVL
GGEIGPETSIFDAGSSLIAMIOIVKLSIDVGEAVEAFAAMDYFAPAPALFRLHG
EAVAPASAVDTLLRRVETGERDALRAELPLLRDTHHEPVGAAADGRVTLTGASGF
IGHVLDRLLAARPLVYSTGDDPYLTGATRFGRQADYARVAISYAEIERVD
RRGPPVDVAVHCGYQVNHLLPLDSHSGSVNTALVRAAALAGARSFAFLASAG
ADRLPLSAATLAVGDYRSKLSIEYVNTLVLCASVSHRPGVLVGHPRDRHL
KDMFTALLETARVGMPLSGHVPVCDVGLADLTLGRDPANPGTADGASRTPSR
SAVVHRTVALDELRLHRTGLTEADVLAPAMFERVDRDGGVPLLAAMQALSGGW
PSAREVDHDIILRLIGTPPTPDAGDRPRTCTTSRGPERRRADTLGTBP"
misc_feature
3576..3596
/note="Region: adenylation domain A1"
3708..3746
/note="Region: adenylation domain A2"
3939..3986
/note="Region: adenylation domain A3"
4089..4100
/note="Region: adenylation domain A4"
4350..4370
/note="Region: adenylation domain A5"
4494..4538
/note="Region: adenylation domain A6"
4593..4610
/note="Region: adenylation domain A7"
4647..4706
/note="Region: adenylation domain A8"
4833..4853
/note="Region: adenylation domain A9"
4893..4910
/note="Region: adenylation domain A10"
5043..5075
/note="Region: chiolation domain"
5343..5375
/note="Region: NADH binding site"
ORIGIN
Alignment Scores:
Pred. No.: 1,12e-143 Length: 14159
Score: 3127.50 Matches: 618
Percent Similarity: 91.40% Conservative: 9
Best Local Similarity: 90.09% Mismatches: 42
Query Match: 87.53% Indels: 17
DB: 1 Gaps: 3
US-10-089-514-2 (1-686) x AF262220 (1-14159)
QY 1 MetArgThrIleuLeuIleAspAsnTyraSPserPheThrHisAsnLeuphegInTyrlle 20
DB 8153 ATGGGACGCTTCGATCGACACTACGACTCGTTCCACGAAACCTGTTCCAGTACATC 8212
QY 21 GlyGluAlaThrGlyGlnProProValValaProAsnAspAlaAspTTPserAryGlu 40
DB 8213 GGGAGGCGCACCGGGCGCCCC---GTGGTGGCCCAACGACGCGGCTGTGGCGGCTG 8269
QY 41 ProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArgGlu 60

DB 8270 CCCCTCGAGAGACTTCACACCGATCCGTCGTCCCGGGCCCGGCGACCCGACCGGAA 8329
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleuGlyVal 80
DB 8330 CGGAGCTTCGGATCAGCCCGCGGCGATCCAGACAGCGGCGCTGCCCTCCGCGCTC 8389
QY 81 CysLeuGlyVHisGlnGlyIleAlaGlnLeuphegIlyThrValGlyLeuAlaProGlu 100
DB 8390 TGGCTCGGCGCACGAGGCGATCCGCCAG-----CTTCGCGCGGAA 8428
QY 101 PromethiGlyArgValSerGlyValaArgHisThrGlyGluAspValaPheArgGlyLeu 120
DB 8429 CCCATCGACGCGCGGCTCTCCAGAGCGCGCACCGGAGAGACTTCCGCGGCGCTC 8488
QY 121 ProSerProPheThrAlaValArgTyHisSerLeuAlaAlaThrAspLeuProAspGlu 140
DB 8489 CCTCGCGCTTCACCGCGGTGCTACCTCCCTGCGCGCACGACTCCCGACGAG 8548
QY 141 LeuGluProLeuAlaIleTrpSerAspAspGlyValaIleMetGlyLeuArgHisArgGlyLeu 160
DB 8549 CTGGAACCTCTCGCTCGGAGCGACGAGCGGCTGCTCATGGGCTGCGGACCGCGAGAG 8608
QY 161 ProLeuTrpGlyValaGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIle 180
DB 8609 CGCGTATGGCGGTCCAGTTCCACCGGAGTCCATCGGACGACGACTTCCGCGCGGAGATC 8668
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisGlnAlaArgArgHisGlyAlaAsp 200
DB 8669 ATGGCCAACTTCGCGCACTCCCTCGCCACACCGGCGACGTCGACCGCGCGAC 8728
QY 201 SerProTyGlyLeuHisValaArgArgValaAspValaLeuProAspAlaGluGluValaArg 220
DB 8729 TGGGCGTACGAACCTCCAGCTGCGCGGCTCGACGTCCTCCCGACCGCGAAGAGTACCC 8788
QY 221 ArgGlyCysLeuProGlyGlyGlyThrThrPheTrpLeuAspSerSerValIleuGlu 240
DB 8789 CGGCGTGGCTCGCGCGCGGCGGCGGCGGCGGCTGAGACGAGCTCCGTCCTCGAA 8848
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrlleuThr 260
DB 8849 GGGCGCTCGCGCTTCCTTCCTCGCGGACGACCGGCGCGCTCGCGGACGTCCTCACCC 8908
QY 261 TyrArgValAlaAspGlyValaValaSerValaArgGlySerAspGlyThrThrArgTrpThr 280
DB 8909 TACCGCGTGGCGGCGGCGGCTGCTCTCGCGGCTCGGACGCGCACGACCGCGGAC 8968
QY 281 ArgArgProPhePheAsnTyrlleuGluGluGluGluArgArgValaProValaAla 300
DB 8969 GCGGCGACCTCTTCAGTACCTCGAGGAGGAGCTGAAACCGCGCGGCGGTCCGTCGCC 9028
QY 301 ProGluLeuProPheGluPheAsnLeuGlyTyrlleValaGlyTyrlleuGlyTyrlleuLeu 320
DB 9029 CCCGACCTGCGCTTCAGTTCAACTCGGCTACGTCGCTACCTCGGCTACGCGTGAAG 9088
QY 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 9089 GCGGAGACCAACCGGACACCCGAGTACCGGCGCGGCGGACCGGACCGCGCTTCCTTC 9148
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlyGlyCysCysTyrlleuLeuAlaLeuAsp 360
DB 9149 GCGGACCGGCGCATGCGCTTCGACCAACGAGAAAGTGCTGCTGCTGCGGCTCGAC 9208
QY 361 ArgArgGlyVHisAspAspGlyAlaArgAlaTrpLeuArgGlyThrAlaGluThrLeuThr 380
DB 9209 CGCGGAGGCGACGACGAGCGCGCGCTGCTGCTGCGGAGACGCGCGACCGCTCACCC 9268
QY 381 GlyLeuAlaValaArgAlaProAlaGluProThrProAlaMetValaPheGlyIleProGlu 400
DB 9269 GGCCTGAGCGTCCGCGCTCGGCGGCGGAGCGGACCGCGCATGCTTCGAGGTCGCCGAG 9328
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspValaAspAlaTyrlleuLeu 420

Db 9329 GCGGCGGCGCTTCGCGCCCTGCTGCGGACGCGACGAGAGAGCGCC----- 9379
Qy 421 Arg1LeuSpGluCysLeuGlu1LeuArgAsnGlyGluSerThrGlyLeuThr 440
Db 9380 -----TCGGCGCTCCGCAACGCGGAGTGTGACGAGATCTGCTGACC 9421
Qy 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuThrSerAlaLeuArgAla 460
Db 9422 AACATGCTCACCGCGCGGACGAGCGACCGCCCTGCGCTTACTCTCGGCGTCCGCCG 9481
Qy 461 IleSerProValProThrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
Db 9482 ATCAGCGCCGCTCCGCTGCGCGCTGCTGAGTGTCCCGAGGCTGTGGTGTCTCAGCGCC 9541
Qy 481 SerProGluLeuPheLeuThrIleGlyAlaAspGlyGlyValGluSerThrProIleLys 500
Db 9542 TCCCGCGAGGCTTCCTCAGATCGCGCGACGCGCGCTCGAGTCCAGGCCATCAG 9601
Qy 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
Db 9602 GGGACCGCGCCCGCGCGGCGACCGCGGAGAGAGACGCGCTCCGCGCGACTGCGCC 9661
Qy 521 GlyArgGluLeuAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
Db 9662 GCGCGGAGAGAGACCGCGCGGAGAACCTGATGATCGTACCTGCTCGGCAACGACTTC 9721
Qy 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTy 560
Db 9722 AACAGCGCTGCGCGAGTCCGCTCCGCTCAGCTGCGCGCTCTTCGAGTGGAGACCTC 9781
Qy 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
Db 9782 GCGCCCGTGCACCACTGCTGTGACCACTCCGGGAGAGCGCTGCGCGCGGCGACGACCC 9841
Qy 581 AlaAlaCysValAlaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProGlyLysArg 600
Db 9842 GCGCGCTGCTACGCGCGCTTCCTCCGCGGCGTCCATGACCGCGCGCGCCAGAGAGCA 9901
Qy 601 ThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrsSerGlyAlaLeu 620
Db 9902 CCCATGAGATCATCGACCCCTCGAGAGAGCGCGCGCGCTTACCCGCGCGCGCTC 9961
Qy 621 GlyThrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
Db 9962 GAGTGTTCGCTCCGCGCGCGCGCGCGCGGACCTGACATCGCTCATCCGACCATGTGCTG 10021
Qy 641 AlaAspGlyGluAlaGluPheGlyValGlyAlaIleValSerLeuSerAspGlnGlu 660
Db 10022 GCGGACGCGCGCGCGCGGCTGCGGCTGCGCGGCGCATCGTCTCCGACGAGAG 10081
Qy 661 GluGluPheThrGluThrValValValAlaArgAlaMetValThrAlaLeuAspGlySer 680
Db 10082 GAGAGATTCAGGACGCGCTGCTCAGGCGCGCGCGCATGTGTACCGCGCTCGACGAGAG 10141
Qy 681 AlaValAlaGluAlaArg 686
Db 10142 GCAGTGCGGCGCGCGCA 10159

RESULT 6
ARI98361
LOCUS ARI98361 2220 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 14 from patent US 6352839.
ACCESSION ARI98361
VERSION ARI98361.1 GI:20248210
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2220)
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and
Dutric-Roseet,G.
TITLE Streptogramins for preparing same by mutasynthesis

JOURNAL Patent: US 6352839-A 14 05-MAR-2002;
FEATURES Location/Qualifiers
source 1..2220
/organism="unknown"
/mol_type="unassigned DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,56e-80 Length: 2220
Score: 1816.00 Matches: 385
Percent Similarity: 64.03% Conservative: 76
Best Local Similarity: 53.47% Mismatches: 207
Query Match: 50.83% Indels: 52
DB: 6 Gaps: 11
US-10-089-514-2 (1-686) x ARI98361 (1-2220)
Qy 1 MetArgThrLeuLeuIleAspAsnTyrsAspSerPheThrHisAsnLeuPheGlnTyIle 20
Db 70 GTGCGAACCTCTGCTGATTCGACCACTACGACTCGTTACCTTCCAGATGCTG 129
Qy 21 GlyGluAlaThrGlyGlnProProValValValProAsnAsp-----AlaAspTPSer 38
Db 130 GCCGAGATGAAACGCGCGCGCTCGCTGCTGCGCAGACGACGACCGGACCTGGCAG 189
Qy 39 ArgLeuProValGluAspPheAspAlaIleValAlaSerProGlyProGlySerProAsp 58
Db 190 GCCCTGCGCGCGCGGCTTCGACCAAGTGTGCTGTCTACCGCGCGCGCGCGCGCGCC 249
Qy 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78
Db 250 ACCGACACCGACCTGAGGCTCAGCGCGCGCGGATATCAAGAAAGGACCTGCGCTGCTC 309
Qy 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98
Db 310 GGGGTGCTGCTGGGCGACGAGCGCTGTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCA 369
Qy 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArg 118
Db 370 CCGGAACCTTTTCACGCGCGCGCACACGACGACATCCGCGCACGAGGAGGCGCTTCCGCG 429
Qy 119 GlyLeuProSerProPheThrAlaValArgTyrsSerLeuAlaIleThrAspLeuPro 138
Db 430 AACATCCCTCCCGCTGACCGGTGCTCCGCTACCTGCTGACCTCCGCGCACTGCGCC 489
Qy 139 AspGluLeuGluProLeuAlaTPSerAspAspGlyValValMetGlyLeuArgHisArg 158
Db 490 GCGGACCTGCGCGCGCACCGCGCACCGCGCGACGCGGACGCTGCGCGCGCGCGCGCG 549
Qy 159 GluLysProLeuThrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178
Db 550 CACCTGCGCGCGCTTCGCGGTGCGATTCACCGCGCATTCATCGACGCGCAACGCGCAC 609
Qy 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193
Db 610 CGGATCTGCGCACTTCGCGACCTGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 669
Qy 194 AlaArgArgHisGlyAlaAspSerPro----- 202
Db 670 ACCGAAGCATACCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
Qy 203 -----TyrGluLeuHisValAlaArgValAlaAspValLeuProAsp 215
Db 730 GCGTCCGCGCGGTGCGGAGTACCGCGTGTGCTGCGGAGTGTCCGTGCGCGCGCGCG 789
Qy 216 AlaGluGluValAlaArgArgGlyCysLeuProGlyGluGlyThrThrPheThrLeuAspSer 235
Db 790 GCGGACGCGCGCTTACCGCGCTGTCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCG 849
Qy 236 SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAlaAspArgGlyProLeu 255
Db 850 AGCGCGTGTGAGGCGCGCGCTGCGCGCGCTTACCTTCTGCGGCGCGCGCGCGCGCGCG 909

[illegible]

	FEATURES
Db	1942 GACGGGCTCCGAAGAGGCCCGCGGGGTACTCGGCGCGCTGACTTCGCCTC 2001
Oy	626 SerGIYAlAalaaPLeSerIleValIIeArgThrIIeValLeuAlaaSPglYInAla 645
Db	2002 AGCGGCGCGCGCACCTCAGCATGTCATCGCACCATCTGTCGCCACCAGAGAGGCCGCC 2061
Oy	646 GluPhcglYValGlYglYAlaIIeValSerLeuSerAspClnlguLtuPhmTrgLu 665
Db	2062 ACCATCGCGCGCGCGCGCGCTGTCGCCCCCTCTCCACCCGACGACGAGTCCCGCAA 2121
Oy	666 ThrValValLYsaIIaAcGAlaMeValThrcAlaLeuaspOlySerIIaValaIGlYAla 685
Db	2122 ATGCTCTCAAGGCGCAGACCACCTCTCGCCGCTTG--CGCCAGGCACAGCGGGCGCC 2178
RESULT 7	
SPPU60417	
LOCUS	SPU60417 4740 bp DNA linear BCF 07-MAR-1997
DEFINITION	Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine precursor biosynthesis (papa, papC, papB, papM) genes, complete cds.
ACCESSION	U60417 GI:1575335
VERSION	U60417
KEYWORDS	
SOURCE	Streptomyces pristinaespiralis
ORGANISM	Streptomyces pristinaespiralis
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 4740)
AUTHORS	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Zagorec, M., Schlemmiger, J., Biech, D., Blanche, F., Debussche, L., Crouzet, J. and Thibaut, D. Identification and analysis of genes from Streptomyces pristinaespiralis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I MoJ. Microbiol. 23 (2), 191-202 (1997) 97197164 9044253
JOURNAL	2 (bases 1 to 4740)
MEDLINE	Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Schlemmiger, J., Biech, D., Blanche, F., Debussche, L., Crouzet, J. and Thibaut, D. Direct Submission Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc Rorer, 13 quat Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403, France
REFERENCES	
TITLE	
FEATURES	Location/Qualifiers 1..4740 /organism="Streptomyces pristinaespiralis" /mol_type="genomic DNA" /db_xref="taxon:38300" 68..2227 /gene="papa" 68..2227 /gene="papA" function="p-aminobenzoate synthase" /codon_start=1 /transl_table=11 /product="papa" /protein_id="AAC4866.1" /db_xref="GI:1575336" translation="MRTVRTLLIDNYSFTYNLFQMLAEVNGAPLVRENDTRTWOA LAGDEPDNVVVSPGPHPATDIDLGLSRITETWDLPLGVLCIGHQALCLGAAVAH APEPEFRSTDIRHDGGELPANIPSEPLTVRVYSHLRQLPADLATTAADOLMAV AHRLPRFGVOFHPRESISBEHGHRMLANPDLSIRAAGRPHPTERIIPAPAPAPAP AAPRASPVEGERLAVREIVACYPPDAADAFALFADAPARFMILDSRVGPFL GARPGLEGOITYDVADRAVRKVDGSGETRRGTLPDHHEHLAKALPATGPREF NLGVALLGYETKADSGEDBAHRELPGAEMADMALLDHEOGRAMLLASSTRP ATAFAERMLTDAARTLATATAPRPFTLLPDDQLPALDVYRHSLRYEELVEECRL ITGOETEVCILTMMRLVPRGIDPLTVBALRTVSPAAYAIOPGATVYLSSPERFL RIQADGWASKESKPICTRPRGAPAOADAOKASIAAEKORSENIMLVIDLVNDLGQC DIGSVHYGCLFEVETVATTHQULYSTRGRLAADVSPPRAVARAFPGSGNTGAPKXTM OFIDRELEKGRGVYSGALGFALSAGADSIIVTRIIVTEEAATTIVGGAIVAVALSPD DEVEMLEKMQDTTLAARQAHAAGATSDRELAAGTSR"
CDS	

Db 1430 TCCGGGCGCAACGGCCGGGAACCCGAGGTATCCGCAACGACGCCGCTGGCGCGC 1489
Qy LeuProValGluAsePheAAspAlaIleValSerProGlySerProAAspArg 59
Db 1490 GGTCTGCTCGACGGCTTCGACACGCTGCTCTCCCGGGGCGGGCACCCGACCGC 1549
Qy 60 GluAAspPheGlyIleSerArgArgAlaIleThrAsePheGlyLeuProValLeuGly 79
Db 1550 CCGGCGCACTTCGCGCTGTGGCGCCGATGCGCGAGAGGCGCGCTGCTGCTCGGC 1609
Qy 80 ValCysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
Db 1610 GTCTGCTGGGCGCCACGAGGATGCGCTCGCCACGCGCGCGGCTGGGCGCGCC 1669
Qy 100 GluProMetHiseGlyArgValSerGluValArgHiseThrGlyGluAAspValPheArgGly 119
Db 1670 GAGCGCCGTCACGGCGCCACCTCGCGGTACGACACGACGACGCGCGGCTTCGAGAGG 1729
Qy 120 LeuProSerProPheThrAlaValArgTyrHiseSerLeuAlaIleThrAAspLeuProAsp 139
Db 1730 CTGCGCGACGCGCTGGAGGTGCTGCGGTACACTCCCTCGCGTGAAGCACTCCGCGC 1789
Qy 140 GluLeuGluProLeuAlaTyrPseAAspArgGlyValValMetGlyLeuArgHiseArgGlu 159
Db 1790 GACTGGAGGCCACCGCGCTGCTGCGAGAGACGGGGTCTGATGGCGCTGCGCCACCGCACG 1849
Qy 160 LysProLeuThrGlyValGlnPheHiseProGluSerIleGlySerAAspPheGlyArgGlu 179
Db 1850 CTGCGCTGTGGGAGAGTGCAGTTCCACCGGAGTCGATCGGACCCGAGCGCGCACCG 1909
Qy 180 IleMetAlaAAspPheAAspLeuAlaLeuAlaHiseIleSarGAlaArgArg----- 196
Db 1910 CTGCTGCGCAACTTCGCGACCTCACCGACCGCCACCGCGCAACGCGCCACGCGCGCG 1969
Qy 197 -----HiseGly----- 200
Db 1970 GCGGGAACAGCGACGCTCCGCGCCCGCGCGCGCGAGAGACGAAGCGCACACCGCG 2029
Qy 201 SerProTyrGluLeuHiseVal-----ArgArgValAAspValLeuProAAspAlaGluGlu 218
Db 2030 ACCCGACGGGCGCTCGCGGTATCGCAAAAGTCCGTGCCACGCGCTGGAGCGCGAGGTC 2089
Qy 219 Val-----ArgArgGlyCysLeuProGlyGluGlyThrThrPheThrPleu 233
Db 2090 GCGTTGACCTCGCTTCGCGACCGCGACCGACCGC-----TTCGTGCTC 2134
Qy 234 AAspSerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAAspAAspArgGly 253
Db 2135 GACAGAGAGCGGTCCGCGGGGCGAGCTGGGCGACGCTCCATGATGGCGGACGCTCAGGT 2194
Qy 254 ProLeuAlaGluTyrLeuThrTyrArgValAlaAAspGlyValIleSerValArgGlySer 273
Db 2195 CCCCTGCGCGGACCGCAAGCGGACGTGCACGCGGAAACGCTCACGCGTGAGA---GCC 2251
Qy 274 AspGlyThrThrThrArgThrArgArgProPheAAspThrLeuGluGluGluGluGlu 293
Db 2252 GACGGCGCCAGCGACGCGTCGAGAGCGCTTCCTGACCTGCGTGGAGAAAGCATGGCG 2311
Qy 294 ArgArgArgValProValAlaProGluLeuProPheGluPheAAspLeuGlyTyrValGly 313
Db 2312 GGGCTCGGACCGAGGTG---CCGAACCTTCCTTGGCGTTCGCGCTCGGCTGGGTGCG 2368
Qy 314 TyrLeuGlyTyrGluLeuLeuValGluThrThrGlyAAspProAlaHiseArgSerProHise 333
Db 2369 TGCCTGGGTACGAGTTGAAGCGCGAGTGCAGCGGACCGCGGACCGCTCGCGCGAT 2428
Qy 334 ProAAspAlaAAspPheAAspPheAlaAAspArgAlaIleAlaAAspHiseGlnGluGlyCys 353
Db 2429 CCGGAGCGCGCTGCTGCTTCGCGACCGGCGCTGCTGTTGAGCAACCGCACCGCGAC 2488
Qy 354 CysTyrLeuLeuAlaLeuAAspArgArgGlyHiseAAspAAspGlyAlaArgAlaTyrPleuArg 373
Db 2489 ACCTACCTGCTGCGCTGCTGAGAGAGACGCGGAGCGGCGCGCTGCTGCGG 2548

Qy 374 GluThrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrProAla 393
Db 2549 GCGGCTCGCGCACCGCTGAGCGCGGTGCGGGCGG-----GAGCGCGAGCGGTGC 2599
Qy 394 MetValPheGlyIleProGluAla---AlaAlaGlyPheGlyProLeuAlaArgAlaArg 412
Db 2600 -----CCGAGGCGCGCGGTGTGACAGACGAGGTCCGGG---GAGCTGCGC 2641
Qy 413 HiseAAspIleAAspAlaTyrLeuLeuValArgGlyCysLeuLeuGlnIleArgAAspGly 432
Db 2642 CACGACCGGAGCGGCTACCTGAGCTGATGACGCTGCGCAGAGAGATAGCGCGCGG 2701
Qy 433 GluSerTyrGluIleCysLeuThrThrAAspMetValThrAlaProThrGluAlaThrAlaLeu 452
Db 2702 GAGACTTACGAGGTCTGCTGACCAACATGCGGAGCGGAGACCGACCTCACCGCGTGG 2761
Qy 453 ProLeuTyrSerAlaLeuAAspAlaIleSerProValProTyrGlyAlaLeuLeuGluPhe 472
Db 2762 GCGGCTACCGCGCGCTGCGCGGAGTGAAGCGCGCGCGCGCTTCGCGCTGCGACTTC 2821
Qy 473 ProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAAspGly 492
Db 2822 GGTCCATGCGCGCTGCTCAGACGCTCTCGGAGCGGTTCTTGCGCATTCGACCGGAC 2881
Qy 493 GlyValGluSerTyrSerProIleGlyGlyThrArgProArgGlyGlyThrAlaGluGluAAsp 512
Db 2882 CCGATGAGTCCAAAGCGGATCAAGGCGACGCGCGCCACCGCGCCACCGCGAGAGAC 2941
Qy 513 GluAArgLeuAAspAlaAAspLeuAlaGlyArgGlyLysAAspArgAlaGluAAspMetIle 532
Db 2942 GCGGCGCTGTACGCTGCTGCGCGCTGCGCGAGAGACCGCGCGGACCTGATGATC 3001
Qy 533 ValAAspLeuValArgAAspAAspLeuAAspSerValCysAlaIleGlySerValHiseValPro 552
Db 3002 GTGACCTGCTGCGCGACGACCTGCGGCGGTGCGGAGTGGCTCGGTCTGCGCGAC 3061
Qy 553 ArgLeuPheGluValGluThrTyrAlaProValHiseGlnLeuValSerThrIleArgGly 572
Db 3062 CCGGTGTTCCAGGTCCAGACGTACGAGCGGACCGCTGCACACTGTGTCAGTCCGACGCG 3121
Qy 573 ArgLeuAAspProGlyThrSerThrAlaAlaCysValArgAlaAAspPheProGlyGlySer 592
Db 3122 CCGCTCGGAGACACAGACCGCGGTGCGCGGTGCGGCGGCTTCGCGCGGCGGCTG 3181
Qy 593 MetThrGlyAlaProLysLysArgThrMetGluIleIleAAspArgLeuGluGluGlyPro 612
Db 3182 ATGACCGGGGCGCGAAGATCCGACCATGAGATGATGACCGGCTGGAAAGCGGGCG 3241
Qy 613 ArgGlyValTyrSerGlyValaLeuGlyTyrPheAlaLeuSerGlyAlaAAspLeuSer 632
Db 3242 CCGGTGTCTACTCGCGCGCATCGGCTACTTCCTCCACGCGCGCGGTAGACCTGTCC 3301
Qy 633 IleValIleArgThrIleValLeuAlaAAspGlyGlnAlaGluPheGlyValGlyVala 652
Db 3302 ATGTGATCCGACCGGTGTGCTCAAGCGGCGGAGGCTCGCTACGCGCTCGCGCGCG 3361
Qy 653 IleValSerLeuSerAAspGlnGluGluGluPheThrGluThrValValLysAlaAAspAla 672
Db 3362 GTCATCGCGCTCTCGACCGCGCGCGACGAGTTGAGAGAAACGCGGTCAAGCGCGCG 3421
Qy 673 MetValThrAlaLeuAAsp 678
Db 3422 CTGCTGCTGCTCTCGAC 3439

RESULT 9
SGR300302
LOCUS 39314 bp DNA linear BCT 20-MAR-2002
DEFINITION Streptomyces grieseus partial ORF1, canA gene, canC gene, canF gene,
canT gene, canB gene and canR gene.
ACCESSION AJ300302
VERSION AJ300302.1 GI:12231153
KEYWORDS ABC-transporter; aminotransferase; can RB gene; canA gene; canC

gene; canf gene; canB gene; canT gene; Cho-like protein;
cytochrome P-450; ferredoxine; PABA synthase; thioesterase.
SOURCE
Streptomyces griseus
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyicinae; Streptomycetaceae; Streptomyces.
REFERENCE
1
Campelo, A.B.
Thesis (2000) Department of Microbiologia, Universidad de Leon,
Leon, Spain
REFERENCE
2
Campelo, A.B. and Gil, J.A.
The candididin gene cluster from Streptomyces griseus IMRU 3570
JOURNAL
Microbiology (Reading, Engl.) 148 (Pt 1), 51-59 (2002)
MEDLINE
21642576
PubMed
11782498
REFERENCE
3 (bases 1 to 39314)
Gil, J.A.
Direct Submission
Submitted (14-NOV-2000) Gil J.A., Microbiologia, Universidad de
Leon, Campus de Vegazana s/n, 24071, SPAIN
FEATURES
source
1.39314
/organism="Streptomyces griseus"
/mol_type="genomic DNA"
/strain="IMRU 3570"
/db_xref="taxon:1911"
complement(<1..1785)
/note="ORF1"
/codon_start=1
/transl_table=11
/product="Cho-like protein"
/protein_id="CAC22112.1"
/db_xref="GI:12231154"
/db_xref="UniProt/TREMBL:Q9EWC6"
/translation="MLERTETIATVDALGATTEGRSSIVLTGPGIGRSALQRL
SCADRDVRAVIRAHAAPEODPARGVQQLPETLLGDSPEDARDMDAHAFARH
VLADDAAPRGADQAIATEAVLHGLSLANYSADSRLLILVDILQMSVPSIRMTY
LAKRLHRAVVCALRGDPSRSHHTLVREIRPAOTLRPASLSLTATHELVREHFG
EAGDDEPVQACHEASVGNPLFILTIIWGTFGGRPLRAHAETARLRPSQSLREIRAS
ILRTOPAPRDLAAITAIIGESDAPTLRIALDSTIGTAGALRALGALAAAPRP
RPTHRVRODAESTLTWQERKHDDPAALIVAGSPACVAAQLMAVTRQPMAYD
VIRAAADTALRRGAPPTAGYLCRALDSPAAGVGRLLVELGTAREGPDPLACBRH
IAQMTLLPEPRDAVAALRIIPTALGPALPTAVDILRQAEDLGPADGIGTDRDLA
LRLEARLRHCGHDPHELAASARHLKEIGPEPVDTGGERLTIALLGATTCGLHPA
SEVVRLLAGRVLDREPAATSRIHSTLPLAT"
9525..10583
/gene="canA"
9525..10583
/gene="canA"
/codon_start=1
/transl_table=11
/product="aminotransferase"
/protein_id="CAC22113.1"
/db_xref="GI:12231155"
/db_xref="GOA:Q9EWC5"
/db_xref="UniProt/TREMBL:Q9EWC5"
/translation="MAFTHEVSRPALDGRLEAVSVAVSGWISOGPVVRRPEEAF
ENGVAVHVAACSSGTALTLALRALITGREDETIVEFTWVASAMAVTTGATPVPD
CGDLDIVRIIEKITPTRAVMPVAVVGRCDMAVMDLAQYNLRVEDSAEAG
VRPVDIACPSLANKIITAGEGVCLTDPRLABDLAHIRAFAPRDSFLKAKL
NYRMTMOGAVALAOETERLDEILATREIRIARVDAGIKDLPGITLMPARDVLMYDLR
AERREELRAHLDERGIETRLFFKPMRQRPYLDPMVPTLNAHRSFSDGLYLPHTGLT
ADQEVITGAVRPYRAG"

/db_xref="GOA:Q9EWC4"
/db_xref="UniProt/TREMBL:Q9EWC4"
/translation="MTSPGPTVVDPFPRPTPEPRLPSQYAEHRKQNGLVQTHLPNGR
PIMVTRHEDVAVLTTPRISANPNDEGPVNGETMGVQKQIDGMFGMLSDPEHR
PRKULIPETPRVAVELRPAIRERTYDEIDAMLAGNTADLVNDRLPSPISVIAL
GPSADRODFEESRTTLVAIRSTIDEEBAETROLRLRIINRLIVYKKKRGSDILSR
LSTGLSDSEESGLVLLLLIAGHETTANNIGAGVVTLSHREWIGDDRLVEBLRHS
VADVVALRAVDVDEIAGQTIKRGCIAPLALASANDTEAFCCPAFNPERTERRHVA
FCYGHQCIQGVLVVEMEIAYRKLFERIPELRVAVPEQQLAYKYDGLIFGHLELPIVR
W"
11906..12100
/gene="canP"
11906..12100
/gene="canP"
/codon_start=1
/transl_table=11
/product="ferredoxine"
/protein_id="CAC22115.1"
/db_xref="GI:12231157"
/db_xref="GOA:Q9EWC3"
/translation="MRVTYDSRQCVGAGQCVLNAPVEFDQDDGVVLLRAEPDRDH
EAVRTAGDLCPSASVVLQED"
12223..12993
/gene="canT"
12223..12993
/gene="canT"
/codon_start=1
/transl_table=11
/product="thioesterase"
/protein_id="CAC22116.1"
/db_xref="GI:12231158"
/db_xref="GOA:Q9EWC2"
/db_xref="UniProt/TREMBL:Q9EWC2"
/translation="MTTADDTAARLRLRHPREDAVNLVCPRPAGSASFYHVSAR
FAPGAEVSLQTPGRQDRKEPCVPDLGLADLITQLPLDERPTVFGSMGALALA
FPTGARLRQKGPRTVIAISGRGSPVTRAEVRHTDDGIVAEKRLNGTAGVAGD
EELILMALPALGEGDRALETYTCPDRLACGLTVLTGDDPLTTVEEAERWRDHTTG
PRLVFTGCHFFLTQHLDVNTETIAQALHPRAAPAA"
13190..15361
/gene="pabAB"
13190..15361
/gene="pabAB"
/codon_start=1
/transl_table=11
/product="PABA synthase"
/protein_id="CAC22117.1"
/db_xref="GI:12231159"
/db_xref="GOA:Q9EWC1"
/db_xref="UniProt/TREMBL:Q9EWC1"
/translation="MRTLLVDNYDSFTYVLPHYLSRANGREPEVIRINDDPAMVRGLD
APDNVLSRPGTGPBHPADFGICARIAEGRGLPVGVCIGHQMLAHGARVGRAP
RHGRISAVRHDGTGLPEGLPOPLEYVRHSLAVTELPELELTANSSEDCVLMALHRT
LPLWGVQHPESIGTQDGRILANFRDLTERGRTRHGRAGHGLPPAPARETKAT
TGTPRLRLVIAKSLPTRMDAEYAFSLFRITGHPRLWLDLSRSGELGQLSNMGDASGP
LARTAKADVAGTIVRADGASITVESAFILTWLENDLAGLRTVEBELPPAFALGHWGC
LGYELKAECDGDAARSPDDAVLVEADRALVLDHRTTLYTLALVEDDABAEAAEMV
LAASVTLDAVABREBPCEAPVCTGPELPHDDGYSKILDVCOEYLAAGETAEVC
LTNMAEDDTLTPMAAYRLRVRSPAPPAFLDPGMATVSSPFFRLIDHGMES
KPIKTRIPGAPQPDADLVRLATCEKDRANLMI VDLVRHDLGRCAVGSVAVADPV
RQVYGAIGAYFELTAGVIDSIYIRIVYLSGSLRYGVGAVYALSDPADEFETAVKA
ARLLLDLDPGAPFGREAPGKDLDGEEDDGDADAAPDVLVPG"
20829..21836
/gene="canRA"
20829..21836
/gene="canRA"
/codon_start=1
/transl_table=11
/product="ABC-transporter"
/protein_id="CAC22118.1"
/db_xref="GI:12231160"
/db_xref="GOA:Q9EWC0"

Db 14762 GCCCGCGCTGCTAGCTGCTCCCTGGCCACCTGCGAGAGAGACCGCCCGAAGAACTGATGATC 14821
Qy 533 ValAspLeuValAlrGAsnAspLeuAsnSerValCysAlaIleGlySerValHisValPro 552
Db 14822 GTGCACTGGTCCGCCACGACCTGGGGCGGTGGCGGAGCTCGGCTCGTGGCGGAC 14881
Qy 553 ArgLeuPheGluValGluThrValAlaProValHisGlnLeuValSerThrIleArgGly 572
Db 14882 CCGGTGTTCCAGGTTCGAGAGTACGCGCACCGCTCACGCTGCTCAGTACCTGCACGGCC 14941
Qy 573 ArgLeuAspProGlyIleThrSerThrAlaAlaCysValArgAlaIlePheProGlyGlySer 592
Db 14942 CGGCTGGCGGAGACACGACCGCGGTGGCGGCGGCTCGGGCGGCTTCCCGCGGGGTG 15001
Qy 593 MetThrGlyAlaProIlySlyAlaArgThrMetGluIleIleAspArgLeuGluGluPro 612
Db 15002 ATGACCGGGCGCGCGAAGATCGGCACCATCATATGACCGGCTGAGAGGGCGGGCC 15061
Qy 613 ArgGlyValIlySerGlyAlaLeuGlyIlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSer 632
Db 15062 CGCGGTCTTACTCTCGGCGCGCATCGCTACTCTCTCCCTCACCGCGCGGTAGACTGTCC 15121
Qy 633 IleValIleIleArgThrIleValLeuAlaAspGlyValAlaGluPheGlyValGlyValAla 652
Db 15122 ATGCTGATCCGACGCGGTGGCTGACGCGCGGAGGCTCGCTACGCGCTCGGCGGGGCC 15181
Qy 653 IleValSerLeuSerAspGlnGluGluPheThrGluThrValValIlyValAlaArgAla 672
Db 15182 GTCATGCGCGCTCTCCGACCGGCGCGACGACTTGAGAGAGCGGCGGTCAAGGCGGCCGC 15241
Qy 673 MetValIleThrAlaLeuAsp 678
Db 15242 CTGCTGCGCTCTCTCGAC 15259

RESULT 10
AY310323
LOCUS 138203 bp DNA linear BCT 06-DEC-2003
DEFINITION Streptomyces sp. FR-008 heptane macroide complex synthesis gene
ACCESSION AY310323
VERSION AY310323.1 GI:34766435
KEYWORDS
SOURCE Streptomyces sp. FR-008
ORGANISM Streptomyces sp. FR-008
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 138203)
Chen, S., Huang, X., Zhou, X., Bai, L., He, J., Jeong, K. J., Lee, S. Y. and
Deng, Z.
Organizational and Mutational Analysis of a Complete
FR-008/Candididin Gene Cluster Encoding a Structurally Related
Polyene Complex
Chem. Biol. 10 (11), 1065-1076 (2003)
14652074
2 (bases 1 to 138203)
Chen, S., Huang, X., Zhou, X., He, J., Bai, L., Jeong, K. J., Lee, S. Y. and
Deng, Z.
Direct Submission
Submitted (30-MAY-2003) Bio-X life Science Research Center,
Shanghai Jiaotong University, 1954 Huashan Road, Shanghai 200030,
China

FEATURES
source
1. .138203
/organism="Streptomyces sp. FR-008"
/mol_type="genomic DNA"
/db_xref="taxon:206662"
complement(574. .1950)
/gene="fscO"
complement(574. .1950)
/gene="fscO"
/note="putative FAD-dependent monooxygenase"
/codon_start=1
/transl_table=11

gene
CDS

/product="FscO"
/protein_id="AA082549.1"
/db_xref="GI:34766436"
/translation="MSTCPAPTSIGPFGCPHRRKGRKRIAVPTPCCGVKRGH
CHAKKRVSMRGGSVAIVGSGITAGCAALASRGGAERTVTERDARDLRDGVGIV
HPARVAELRDAGYMDDEMPAPLNRRWSVRDPAEHRALICTOPPPPAVSGSLMS
ELRRVPEPTAIRAGAKVAVEDDAGTVRLADSEPRFDVIGADGRSVRELMY
PGTGAIVAGYIGMRGTSPDVSGLPSDGLDANHITPEPGCMAYRI.PDSGGRLMWL
YTPAPRIDGLHEDLTPTSLPGRINAEITELRALVLEHPPFWAAKLSTPAETTF
IQPIVDLVPRVATDRMVLIGDASVAPRHLDAGSVKALQDATALAAMVAGESWKEI
LEGVHAARGPVTAVGLARRNGSAQVETTPWSAMGAEFPAMMQEONNGSDRSGF
GCHSLKSR"
complement(2264. .3037)
/gene="pabC"
complement(2264. .3037)
/gene="pabC"
/note="putative 4-aminio-4-deoxychorismate lyase"
/codon_start=1
/transl_table=11
/product="PabC"
/protein_id="AA082550.1"
/db_xref="GI:34766437"
/translation="MIEIDGEPAGPEALASALTNVGHFTLLVNGRVRYDLHLER
LIRDCRTEPDALDPDRVKLARRAAPDGRATVAVTVYDPALNIGNTAAADRCILY
TSRPADPFGPLRVRSVVRHRLDPEVGSVGLPTLRLRQRLQRAIGYDVLFTGPDG
ILEGTMVNGVLVDEVWVPGSEVLAGLITRQLRRGHRSHRTGRPRGRLSRGGLRD
QRVVVRPVYTGIDREFPAAHSHVRLAEIYQALPGSPPL"
complement(3150. .3818)
/gene="fscR1"
complement(3150. .3818)
/gene="fscR1"
/note="putative transcriptional activator"
/codon_start=1
/transl_table=11
/product="FscR1"
/protein_id="AA082551.1"
/db_xref="GI:34766438"
/translation="MDPAPAADPAGSPERRAAVLADGDFRAGAYVACLDPSLTIQVNO
EERERFGSSSELCSRFCDLVHPVOQPLMKHOFAPMLDGRKRPATEVIAVDOERTA
STLPNALAVRGRTPEVVAIIVWMAAEEDGADVMAPRKLLSEIDARILLEGIAA
GVSTPLASRLVLSRQGVHYHTGLRKLVKNRAALVSAISMGVLKVTMPKPYVE
DPIK"
complement(4377. .7205)
/gene="fscR11"
complement(4377. .7205)
/gene="fscR11"
/note="putative transcriptional regulator"
/codon_start=1
/transl_table=11
/product="FscR11"
/protein_id="AA082552.1"
/db_xref="GI:34766439"
/translation="MLERTETIALVDAALGAATGSRSLVLTGPIGIGRSALORL
SLIADRDVDVRLRAHAAPMEDQPAFQVYRQIFPTLIGDSPEADADRMDADAAPFAH
VLADDAAPGADQAIATEAVLHGLSLIANVASDSRLILVDDIQMSVPSRLMITY
LAKRLHGRAVAVVCAIRDDBPSRSHLTAREADQOTLRPLSLTATNHELVRHFG
EAGDPEVQACHASVGNPLFLSLIVGTFGRPLAEHARTARLPPSCORLEAS
IKRTOPAPVRDIAAAILIGEOSDAPTLARLAGDSIGAGALRALGALGAADPEP
RIRHSVYDAESTLTMOOREMHDEAAALVIAASPAOVAAQIMAAVVTROPAAV
VIRAAADTLRRGAPDTANGYICRALDPSPAAGVGRGLVLELTABRGFDPLACRH
IAQAMTLLPEPRDAVVAALRISPTALGAPVLAVDLIRQAEDLDPADLTGTDOLA
LRLRLRCHGEDHDEHLAASARLKEIPPEPVDTGGERLTIALIGATATGCHLPA
SEVVLACGVLDREPTASRIHSTPLAITTIAADSVAVDSMLAVERANRGCTP
ANALVHIBQALVHLGRGLAQRBOSEETLIGTEPEAHQGLDPAALMTAIVALESDP
SLAEKYLRRASRIIPQGLTITGILILERTREIRIAGDQALDSHSCGROLEAGCMH
NFVLFPWRPAVGVHRLGDIRTARSLEBEYRARBDQAPVGVGRALRLRLRLEDG
RGIGLIREAVTVLRGSANOLELARAISTLAELGGAGLETBLAREATLATISCGAPGL
AERARAGHRSGSPATEDEVLTPTERRVWGLACRGLTWOETAGALGVSSRAVEKRLT
HAYRLGISGRELIVLSRAEYAGTD"
complement(7210. .10254)
/gene="fscR11T"
complement(7210. .10254)
/gene="fscR11T"

gene
CDS

```
/note="putative transcriptional regulator"
/codon_start=1
/translate=1
/product="fscR11"
/protein_id="AA082553.1"
/db_xref="GI:34766440"
/translation="MYDLHMPLEPLEERRELAASAVIGELSGRAVVTVTGEP
GLGONDLRWMAAAYADTAGIVLSAHTPEHEVRYGVVAQLAEENRALAPRLFLD
EQQGLGELGLNLAIVDRPTLLVVRDTOMLDPASLRLWLEALRLPRVALTSTT
GTALTREMSVGTPLTGLATTELALPTLDSGGATATVLRFGAPGPATPEALERT
RGI.PAVVHDVLDPRFAGYSPRADRIDPLRLAEVVDGATRALSDRLPRVADLR
ALAVCGDLDFPLVCTLAGHSVSESLRALAASGLTTLRDHPVQDVAARVLE
EMPAORAEVYARAAQRAQVAADDOGIADLRLARVGDPMADITLRPFTALGGE
RDLAVAYLARLADDEPLAEDRARIEFOLASVENVTAPTAERLIGLJATPGGA
GRABATDLCIGSDTPRAARHALAGATDSAPAPEPRGPTGA.PAVPEPEASTPT
HIVPHRYACTPGGATGPDGRSAPRLPCTPRANDPSPGPEAGRPPTAGDEDE
AVAREHELVLCVASFRLHDDADLDVLPALPDVAHNPAVAGVMAWEGVAVNRE
EVRLARSAFVPAPATGRPLLVPRLLGCRALLTDDDEAHNHLGALLDSHERAI
SVHILTVRAELHIRHGRPDAAADLAAQAEPLDLHPLFLPYMLLSMTIDLNG
HTRARRETAARPLPLAHSATTAOLIFAVGLARPTDEHPOAREHPRACRWLTHG
CANPAVOPMRSALAAAHNTGDTBEAARIVHEVRRLARWGAAPRIGRAOLSAVTE
ENREVDRAVTTLSAASARTATYTRAVIELAEERENERTVALGPATVLPAPR
PGPHVPSAVARPPGPRAVLTGRITAKAGHDAAWHGTLAVERETAALAAQGLNR
EIAETLAVTTRAVELRLSGYRKIRIRGREERLALVOEAGS"
complement(10298..13315)
/gene="fscR1V"
/complement(10298..13315)
/gene="fscR1V"
/note="putative transcriptional regulator"
/codon_start=1
/translate=1
/product="fscR1V"
/protein_id="AA082554.1"
/db_xref="GI:34766441"
/translation="MWLRWAGDSRGAAPRSGSAGBPSLAPDARESPVPHRSAP
PAPPLVGRHQLDVLRAHADARAPRPLTLDGPAIGTALLRALAEDGLAGT
TLVSCRAVNAVATGSGVRALFGALGTGKRTSPVLVCGARAPLALADGELDA
DPGSTFVLOGLYWLAANLMAADGPIVLDVDMCDERSLRMDITLRADVPLIV
AAHRTGCTGLAPALADLVANHLPASIGLGRPLTAAYAEITANSPEODPLSEVRL
LSTVGSPLFETVRLDLDRSAGLPDGTGARLQAQGVAAVSVGEHOQDWRO
VAVAVLADGEDTTLAALGRSVVHEAVEILRDAGLHPHLESLHDVAAQAVLD
AVASGVAVLRRRAARLSDVGRSPEBIAQLLPCTDMANVSLRLDAQAQERG
ACAAARYLERVEAREKDPDVLRLGKALAEPTDPAASVTLHEASLITDVARAT
AOLGLTCLAAVQSPDGAARLTEALDALDELGPEDPARELRITLTESALLVSGDE
KATLPILRTBESLTPOQDTPAQOQAMLSVLSAEGGASBETGKARARARAGV
PLGWSLPLPTSLALSDENEAEEVLETVLRSGSDTAAYTVLALSTRSLRENG
AVPDAMADQTALEILGOERMGDTATMPHTAVASVLTERRDPRALDALIKRPHLD
RFVWEHMYLMARGLAADGDIDGALOVFGSCGASMAEAGLTNPVLAAPWLETCIL
GEARGEEBGRAAAAGHRLAERWGTALGYALARGAARGATPGALREAVALLAD
SPARSARGALLLGRALVEDGAVRGRHRLRAVGLARCCGVALLARRADLILAG
GRRREVTASPLDMLTGERTVAGLVASGAKRENVASLFTVATVELHLTSYRKLV
ARRGDLTEALREMGATAPRAERRPGHAKRNP"
13522..14898
/gene="fscM1"
13522..14898
/gene="fscM1"
/note="glycosyltransferase responsible for mycosamine
attachment"
/codon_start=1
/translate=1
/product="fscM1"
/protein_id="AA082555.1"
/db_xref="GI:34766442"
/translation="MDSARPIILFVSLPESGLNPLVLVAGELARGVDPDLWFATDEH
RREVEVALDSVSKVSFASLGEVPELSAVWDEVYREYQSRRFAHARVAVOSRR
AARAKYRELEQVDEVPRALVWDCVAGGVDLARGLPYVALPFAVSNVLTSHN
PFGATYPSFVPSNGLPAPMSVROKLTNTLPEKMTLCEFLHPDMAALLRBAARLK
BLCIAPRNATRVDEAAAYVCSSVAELIDPFDIPDVSLVGAFLPLPFAPLDDEVTR
WLDAGSSVVMGFGLTTRLTREVVAALVEVARRMSGTHQFLMWLAPREOQLLEASOL
PDLRVESWVPSQLDVLAHPNVSVFSPHSGNAHGVFGKPOVVRPLMVDCFDQAV
RGDFGSLTLDKPHTVDPDDVDVLDKLTVTSDBAFRTAERLGLAQRAGGAAADL
VTGLPAA"
14953..16011
```

```
CDS
/gene="fscM1"
14953..16011
/gene="fscM1"
/note="putative aminotransferase responsible for
mycosamine biosynthesis"
/codon_start=1
/translate=1
/product="fscM1"
/protein_id="AA082556.1"
/db_xref="GI:34766443"
/translation="MAFTHPALDRELEVSDAVSGWISSOGPYVRRFEFAFA
ENMGVAGVACSGSGATLTLRALNIGDEVIYBFTMVASMAVATGATPVVD
CGDDLIDVTRIEEKITPRTRAMVPHVYGRCDMDVMDLQVYLRYVEDSABNG
VHPVGGIAFESLPANKITTAGGGVCLTDPLRABDLARMAFTRRDSFLHKILAY
NYRMTMQAVALAQTERLDELATRRREIARIDAGLKLPGITLMPADVLMYDLR
AERRELRHLDRGLETLFPKPSRQPGYLDPPWPTLNHAFSBDGLYLPHTGLT
AADQEVITGAVDPFYRAG"

Alignment Scores:
Pred. No.: 6,52e-76 Length: 138203
Score: 1766.00 Matches: 381
Percent Similarity: 63.93% Conservative: 71
Best Local Similarity: 53.89% Mismatches: 209
Query Match: 49.43% Indels: 46
DB: 1 Gaps: 13

US-10-089-514-2 (1-686) x AY310323 (1-138203)
QY 1 MetArgThrLeuLeuIleAspAsnYrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 18610 ATCGGACCCCTTCTGTCGACAACTACGACTGTTCACTTCCACTACCTC 18669
QY 21 GlyAlaIleThrGlyGlnProProValValValProAsn---AspAlaAspTyrSerArg 39
Db 18670 TCCCGGCGCAACGCGCGGACCGGAGTCTCCGCAACGACGACCGCGCTGCGCGCC 18729
QY 40 LeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArg 59
Db 18730 GGTCTGCTGACCGCTTCCACAACTGCTCTCCCGGCGCGGTACCCACACCGC 18789
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGly 79
Db 18790 CCGCGGACCTTCCGCTGCTGCGCGCGGACCGGACCGGACCGCGCTGCGCTCGCG 18849
QY 80 ValCysLeuGlyHisGlnGlyIleValGlnLeuPheGlyIleThrValGlyLeuAlaPro 99
Db 18850 GTCTGCTGCGGCAACGAGGCACTGCGCGGACCGGCGCGGCGCGGCGCGCGCC 18909
QY 100 GluProMetHisGlyValArgValSerGluValValGlyIleThrGlyGluAspValPheArgGly 119
Db 18910 GAGCCCGCTCACGCGCGGCACTTCCGCGGCTACGAGCGGACCGGCGCTTCCGAGGG 18969
QY 120 LeuProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAsp 139
Db 18970 CTGCCCGACCGCTGAGGCTGCTGCGGTACCTCTCCGCTGACGGAACCTGCGCGCG 19029
QY 140 GluLeuGluProLeuAlaIleTyrSerAspAspGlyValValMetGlyLeuArgHisArgGlu 159
Db 19030 GAGCTGAGCGCACCGCTGCTGCGAGCGAGGCTCTGATGCGCTGCGCACCGCACG 19089
QY 160 LysProLeuTyrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGlu 179
Db 19090 CTGCCGCTTGGGGGGTGCAGTTTCCACCCGAGTGCATGCGGACCCAGGACCGCACCG 19149
QY 180 IleMetAlaAspPheArgAspLeuAlaLeuAlaHisHisArgAlaArg----- 195
Db 19150 CTGCTGCGCAACTTCCGCACTTCAACCGAGCCGACCGGCGGCGCGCGCGCGCG 19209
QY 196 ---ArgHisGly-----AlaAsp 200
Db 19210 GCGGGGCAAGGACGCTCCCGCGCGCGCGCGCGCGCGGAGACGAGCGGACCGCGCG 19269
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuPro-----AspAlaGlu 217
```

Db 19270 ACCCCACGGGCGCTCCGGGCTC---ATCGCCGAGTCCGCTCCACCGGCTGGGACGCGCGAG 19326
QY 218 GlnVal-----ArgArgGlyCysLeuProGlyGluGlyThrThrPheTrp 232
Db 19327 GTCCGCTTCGACTCGCTGTTCCGACCGGCGACCCACCC-----TTCTGG 19371
QY 233 LeuAspSerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArg 252
Db 19372 CTGACAGAGCGCGCTCCCGGGGGGAGCTGGGCGAGCTCTCCGTGATGGCGACGGCTCA 19431
QY 253 GlyProLeuAlaGluThrLeuThrTyrArgValAlaAspGlyValSerValArgGly 272
Db 19432 GGTCCCTCGCCCGGACCGCCCAAGCGCAGCTGCACGCCGGAACGGTCAACGGTGAGA--- 19488
QY 273 SerAspGlyThrThrThrArgThrArgArgProPhePheSerTrpLeuGluGluIndLeu 292
Db 19489 GCCGACGGCGCCGACGACCGGTCGAGAGCGCCTTCCTGACCTGGCTGGAGAACGACCTG 19548
QY 293 GluArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrVal 312
Db 19549 GCGGGGCTGGCGACCGAGG---CCGGAACCTTCCTTCGCGCTTCGCGCTGGCTGGCTG 19605
QY 313 GlyTyrLeuGlyTyrGluLeuGlyAlaGluThrThrGlyAspProAlaHisArgSerPro 332
Db 19606 GGTGCTGGGCTACGAGGTGAAGCGGAGTCGACGCGGCGACCGCGGCAACGCTCACCC 19665
QY 333 HisProAspAlaAlaPheLeuPheAlaAspArgAlaAlaLeuAspHisGlnGluGly 352
Db 19666 GATCCGACCGCCGTGCTGCTTCGCGACCGGCGCCTGTGCTGACCCACCCACCCCG 19725
QY 353 CysCysTyrLeuLeuAlaLeuAspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeu 372
Db 19726 ACCACTTACTCTGCTGGCGCTGGTGGAGAGACGACGCGGACCGACGCGCTGGCTG 19785
QY 373 ArgGluThrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrPro 392
Db 19786 GCGGCGGCTCCGCGCACCTGAGGCGCATCGCGGCGG-----GAGCGCGAGCGG 19836
QY 393 AlaMetValPheGlyLeuProGluAla---AlaAlaGlyPheGlyProLeuAlaArgAla 411
Db 19837 TGC-----CCCGAGGCGCCCGTGTGACGACGCGGTCCGGG---GAGCTG 19878
QY 412 ArgHisAspLysAspAlaTyrLeuLysArgGlyAspGluCysLeuGlyGluLeuArgAsn 431
Db 19879 CGCCACGACCGGAGACGGCTTACTGAACTGATGACGCTGCGACGAGGATGACCGCC 19938
QY 432 GlyLysSerTyrGluLeuCysLeuThrAsnMetValThrAlaProThrGluAlaThrAla 451
Db 19939 GGGGAGACCTACAGAGCTGCTGACCAACATGCGGCGAGCGGACACGACCTCACCCCG 19998
QY 452 LeuProLeuTyrSerAlaLeuArgAlaLeuSerProValProTyrGlyAlaLeuGlu 471
Db 19999 TGGGCGGCTTACCGCGGCTGCGCCGCGGTGAGCCCGCGTTCGCGCTTCGTCGAC 20058
QY 472 PheProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAsp 491
Db 20059 TTGGGCCCAATGGCGCGTCTCAGAGCTCTCCGAGCGGTTCTTCGCGCATGACCGGAC 20118
QY 492 GlyGlyValGluSerLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGly 511
Db 20119 GGGCGGATGAGTCCAAAGCGGATCAAGGCGGACCGGCGGCGGCGGCGGCGGCGGAG 20178
QY 512 AspGluArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMet 531
Db 20179 GACGCGCGCTCTTACGCGCGCTTGCGCACTGCGAGAGGACCGCGCGGAACTGATG 20238
QY 532 IleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisVal 551
Db 20239 ATGCTGACCTGTGTCCGCGACGACTGGGGCGGTGCGCGGAGGTGCGCTGCTGCTGCC 20298
QY 552 ProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArg 571

Db 20299 GACCCGGTGTCCAGGTGAGAGCATACCGGACCTGTGCACCACTGCTGATCCTCACG 20358
QY 572 GlyArgLeuArgProGlyTyrSerThrAlaAlaCysValArgAlaAlaPheProGlyGly 591
Db 20359 GCGCGGCTCGCGAGGACGACGACCCGGTGGGCGGCTCCGGGCGGCTTCCCGGCGGA 20418
QY 592 SerMetThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeuGluGly 611
Db 20419 TCGATGACGGGGGCGCGGAGATCCGACCATGCAATGACATGACCGGCTGGAGAGCGGG 20478
QY 612 ProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAspLeu 631
Db 20479 CCGCGGAGTGTACTCTCGGCGCCCATCGGCTACTTCTCCCTACCGGCGGCTGAGCTG 20538
QY 632 SerIleValIleArgThrIleValLeuAlaAspGlyGluAlaGluPheGlyValGlyGly 651
Db 20539 TCGATGTATCGGACGGGTGTCTCAAGCGGCGGAGGCTGCTGCTGCGCGG 20598
QY 652 AlaIleValSerLeuSerAspGlnGluGluGluPheThrGluThrValValLysAlaArg 671
Db 20599 GCGGTATGCGGCTCTCCGACCGGCGGCGGACGAGTTGAGAGAGACGGCGGTCAA 20658
QY 672 AlaMetValThrAlaLeuAsp 678
Db 20659 CCGCTGCTCGTCTCTCGAC 20679

RESULT 11
AP006618_17/c
WPCOMMENT
Sequence split into 61 fragments LOCUS AP006618 Accession AP006618
Fragment Name Begin End
AP006618_00 1 110000
AP006618_01 100001 210000
AP006618_02 200001 310000
AP006618_03 300001 410000
AP006618_04 400001 510000
AP006618_05 500001 610000
AP006618_06 600001 710000
AP006618_07 700001 810000
AP006618_08 800001 910000
AP006618_09 900001 1010000
AP006618_10 1000001 1110000
AP006618_11 1100001 1210000
AP006618_12 1200001 1310000
AP006618_13 1300001 1410000
AP006618_14 1400001 1510000
AP006618_15 1500001 1610000
AP006618_16 1600001 1710000
AP006618_17 1700001 1810000
AP006618_18 1800001 1910000
AP006618_19 1900001 2010000
AP006618_20 2000001 2110000
AP006618_21 2100001 2210000
AP006618_22 2200001 2310000
AP006618_23 2300001 2410000
AP006618_24 2400001 2510000
AP006618_25 2500001 2610000
AP006618_26 2600001 2710000
AP006618_27 2700001 2810000
AP006618_28 2800001 2910000
AP006618_29 2900001 3010000
AP006618_30 3000001 3110000
AP006618_31 3100001 3210000
AP006618_32 3200001 3310000
AP006618_33 3300001 3410000
AP006618_34 3400001 3510000
AP006618_35 3500001 3610000
AP006618_36 3600001 3710000
AP006618_37 3700001 3810000
AP006618_38 3800001 3910000
AP006618_39 3900001 4010000
AP006618_40 4000001 4110000
AP006618_41 4100001 4210000

AP00618_42 420001 431000
AP00618_43 430001 441000
AP00618_44 440001 451000
AP00618_45 450001 461000
AP00618_46 460001 471000
AP00618_47 470001 481000
AP00618_48 480001 491000
AP00618_49 490001 501000
AP00618_50 500001 511000
AP00618_51 510001 521000
AP00618_52 520001 531000
AP00618_53 530001 541000
AP00618_54 540001 551000
AP00618_55 550001 561000
AP00618_56 560001 571000
AP00618_57 570001 581000
AP00618_58 580001 591000
AP00618_59 590001 601000
AP00618_60 600001 602125
Continuation 118 of 61 of AP00618 from base 170001 (AP00618 Nocardia farcinica IFM 1)

Alignment Scores:

Pred. No.:	1,25e-69	Length:	110000
Score:	1635.50	Matches:	364
Percent Similarity:	60.40%	Conservative:	60
Best Local Similarity:	51.85%	Mismatches:	241
Query Match:	45.77%	Indels:	37
DB:	1	Gaps:	13

US-10-089-514-2 (1-686) x AP00618_17 (1-110000)

QY 1 MetArgThrIleuLeuIleAspAsnTyRAspSerPheThrHisLeuPheGlnTyR11e 20
DB 54159 ATGCGACGCTGTGTGATGACAACTACGACTCGTTCACCTACCAACCTGACAGCTGATC 54100
QY 21 G1yGluAlaThrGlyGlnProProValValValProAsnAsp---AlaAspTPSerArg 39
DB 54099 AGTGAAGGTGAACGGGGTGCAGACCGACCTGCTGCGCAACAGACAGACCGCGCGCG 54040
QY 40 LeuProValG1uAspPheAspAlaIleValValSerProG1yProG1ySerProAspArg 59
DB 54039 CTGGAACCTCGCCCGCTGACGACCAATCGTGTCTCCCGCGCGCGCGCGCGCGCGCG 53980
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGly 79
DB 53979 ACCCGGACGCTCGGCTGTCGACGCGATCAGCGCGGCTGCCAGCGCTGTCGCG 53920
QY 80 ValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
DB 53919 GTGTGCTGGGGACCAAGGACATCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53860
QY 100 GluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGly 119
DB 53859 GTRACCCGACGCGCTTCTCGACCGGGTCCGACGCGGCGCGCGCGCGCGCGCGCG 53800
QY 120 LeuProSerProPheThrAlaValArgTyHisSerLeuAlaIleThrAsp---LeuPro 138
DB 53799 CTGCCGACGATTCACGCTGCGCTGACCTCGCTGCGCGCGCGCGCGCGCGCGCG 53740
QY 139 AspGluLeuGluProLeuAlaIleTPSerAspAspGlyValValMetGlyLeuArgHisArg 158
DB 53739 CCCGCCCTCGAGGTACTGCTCCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCG 53680
QY 159 GluValProLeuTrpGlyValAlaGlnPheHisProGluSerIleGlySerAspPheGlyArg 178
DB 53679 CACGCGCCGACGAGGTGTGTGAGTTCATCCGAGTCGAGTCCGCGCGCGCGCGCGCG 53620
QY 179 GluIleMetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArg-----193
DB 53619 GCGCTCTGCGCAACTTCGCGCAACCTGACCGCGCGCGCGCGCGCGCGCGCGCGCG 53560
QY 194 -----AlaArgArgHisGlyAlaAspSerPro-----202

DB 53559 CGTCCGCGACACACCGCTGACCGCCGCCGTCGCGCGCGCGCGCGCGCGCGCGCGCG 53500
QY 203 -----Tyrg1LeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
DB 53499 CTGCCGACGAGGTATTCAGCGCGCGCGCTGACAC-----GAGCGCGCTTCGCGCG 53446
QY 221 ---ArgG1CysLeuProGlyGluGlyThrThrPheThrLeuAspSerSerValLeu 239
DB 53445 CTTCACGCGCGCTGCGCG-----ACGCGCTTGGCTCGACAGCGAACACGTCGAG 53395
QY 240 GluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyRLeu 259
DB 53394 CCGGCGCTGACCGGTTCTCTTCTTCCTCGGTGAGCGGAGCGAGACCGTGGCGAGGTGTG 53335
QY 260 ThrTyRArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArg 279
DB 53334 CGTTACCGGCTCGCTGCGACGAGTGTGAGAGACCGCGGAAACGCGAGCG---53278
QY 280 ThrArgArgProPhePheAsnTyRLeuGluGluGlnLeuGluArgArgValProVal 299
DB 53277 GTCCCGCGGACGCTGCTGGGTATCTGTCCACCACTCGCGAGCGACCTGAGCTG 53218
QY 300 AlaProGluLeuProPheGluPheAsnLeuGlyTyRValGlyTyRLeuGlyTyRLeu 319
DB 53217 ---CCGCACTGCGCTTGCATCTTCGCGCGGATTCGCTGCTACCTCGGCTACGAGCTC 53161
QY 320 LysAlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeu 339
DB 53160 AAGCGCGACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 53101
QY 340 PheAlaAspArgAlaIleAlaLeuAspHisGlnGlyCysSerTyRLeuAlaLeu 359
DB 53100 TTGCGCGACGCGGTGATCTGTGTCGACATGTGCGCGGCGACACCACTGTGTGCGCTC 53041
QY 360 AspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGlyLuthraLagLuthraLeu 379
DB 53040 -----ACGACGACGACACACCGCGCGCGATCGCTGCTCGGAGACCGCGCGCGCTG 52987
QY 380 ThrGlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
DB 52986 GAACGCTG-----CCCGGCTGGCGAATCCGCGCGAATTG---GCGATCGAG 52942
QY 400 GluAlaAlaAlaGlyPheGlyProLeuAlaIleArgAlaHisAspLysAspAlaTyRLeu 419
DB 52941 TCCGCGCGCGCTCGGCTCGCGCGCTG-----CTGAACCGGCGCGCGCGCGCTACCTC 52888
QY 420 LysArgIleAspGluCysLeuLysGluIleAspAsnGlyGlySerTyRLeuIleCysLeu 439
DB 52887 GCGACATCGCGCGCTGCTTCAGATCTGCGCGCGCGAGTCTTCAAGATCTGCTG 52828
QY 440 ThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyRSerAlaLeuArg 459
DB 52827 ACCGACAGCTCAGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52768
QY 460 AlaIleSerProValProTyrglyAlaLeuLeuGluPheProGluLeuSerValLeuSer 479
DB 52767 CCGTCGAATCCGCGCGCTTATCCGCTTACCTGCGCTTGAACGACTCCAGATCGCTCG 52708
QY 480 AlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerTyRProIle 499
DB 52707 TCTCTACCGGAGCGCTTCTTCAAGATTCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 52648
QY 500 LysGlyThrArgProArgGlyGlyLuthraLagLuuAspGlyArgLuthraLagLeu 519
DB 52647 AAGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52588
QY 520 AlaGlyArgGluLysAspArgAlaGluLeuMetIleValAspLeuValArgAsnAsp 539
DB 52587 GCGGACAGTCCCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 52528
QY 540 LeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThr 559
DB 52527 CTGCGCGGAGTTCGACATTCGCGAGCTGACGATCCGACCTGATGCGCGCGAGTGTG 52468

QY	560	TyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSer	579
Db	52467	TATTCCACCGCTCCACCAAGCTGGTTCAGACGATGGCGGACCCCTGGCCCGGACGTGCAC	52408
QY	580	ThrIleAlaCysValAlaArgAlaAlaPheProGlyGlySerMetThrGlyAlaPolys	599
Db	52407	GCCATCGACTCGCTCGCGGCGCTGTTCCCGCGCGCTCCATGACCGGCGCGGAAGCTG	52348
QY	600	ArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGlyAla	619
Db	52347	CGCAGCATGTGAGATCATCGACAGAGCTGGAAGGCGCGCGCGGGGCTTACTCCGGCACC	52288
QY	620	LeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleVal	639
Db	52287	ATCGGCTTCTCGGACTCGGCGGACCGCGGACCTCAACATCGTCACTCGCACCGCGTG	52228
QY	640	LeuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGln	659
Db	52227	CGGACGACGCGGCGGCTGGCGGATGGGGCGGCGCGCGCGCATCGTCTCGGATTCGCG	52168
QY	660	GluGluGluPheThrGluThrValIleValIleAlaArgAlaMetValThrAlaLeuAspGly	679
Db	52167	GCGAGGAGGTATCCGAGATGTGTGAGAGGCGGCGCGGACCGAGCGCGCGCCACGCGC	52108
QY	680	Seq1a 681	
Db	52107	GCGGCG 52102	
RESULT 12			
LOCUS	AP003593	333550 bp	DNA linear BCT 28-NOV-2001
DEFINITION	Nostoc sp. PCC 7120 DNA, complete genome, section 13/19.		
ACCESSION	AP003593	BA000019	
VERSION	AP003593.1	GI:17132540	
KEYWORDS			
SOURCE			
ORGANISM	Nostoc sp. PCC 7120		
	Nostoc sp. PCC 7120		
	Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.		
REFERENCE			
AUTHORS	1		
	Kaneko, T., Nakamura, Y., Molk, C.P., Kuritz, T., Sasamoto, S.,		
	Watanabe, A., Iriuguchi, M., Ishikawa, A., Kawashima, K., Kimura, T.,		
	Kishida, Y., Kohara, M., Matsunoto, M., Matsuno, A., Muraki, A.,		
	Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M.,		
	Yasuda, M. and Tabata, S.		
	Complete genomic sequence of the filamentous nitrogen-fixing		
	Cyanobacterium Anabaena sp. strain PCC 7120		
JOURNAL	DNA Res.	8 (5), 205-213 (2001)	
MEDLINE	21595285		
PUBMED	11759840		
REFERENCE	2 (bases 1 to 333550)		
AUTHORS	Kaneko, T.		
TITLE	Direct Submission		
JOURNAL	Submitted (02-May-2001) Takakazu Kaneko, Kazusa DNA Research		
	Institute, The First Laboratory for Plant Gene Research, Yana		
	1532-3, Kisarazu, Chiba 292-0812, Japan		
	(E-mail: kaneko@kazusa.or.jp,		
	URL: http://www.kazusa.or.jp/cyanobase/		
	Tel: 81-438-52-3935 (ex. 2338), Fax: 81-438-52-3934)		
FEATURES			
SOURCE	Location/Qualifiers		
	1..333550		
	/organism="Nostoc sp. PCC 7120"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:103690"		
	/note="synonym: Anabaena sp. PCC7120"		
	complement(52..1455)		
gene	/gene="a13408"		
	complement(52..1455)		
CDS	/gene="a13408"		
	/note="ORF_ID:a13408"		
	/ORF_ID:a13408"		
	/codon_start=1		
	/transl_table=1		
	/product="nicotinamide nucleotide transhydrogenase,		

subunit beta"
/protein_id="BAB75107.1"
/db_xref="GI:17132541"
/translation="MSDFLTPTGIQITLYLVAASLFLGLKKLGSPATARNGNVIGAVGM
LLAIAYATLDOHVNLEWILLGLAIGSGIGALAAVKQMTPEPMQGLNGLGGAASA
LVAAVEFRRLLIDAGAPVLDVNIISMLDLVIGVTFPGSGFLAFKLGGLSGSPTFP
LQOPNNILLGGVLAGSAYLITPDSLPVLAVAVASLVLYGMFVIPGGDMPVYS
LINSISGIAAAAAGFVVMNNMLITAGLVASGLILTEINCKMANSLSPEVLSARG
VSAGAGATATDTQTVASIDPEEGAMLTGARSVIVPQIGMAVAQSVBELVQD
LERMVGVDIAIHPVAGMPEGMMNVLLAEANVPYTQLYMEDIINPEQDAVDALVGA
NDVVAPARSSTFASPIGMPILLEVDRAKHTIVIKRGSAGFAGVDNLFYKDXITMLF
GSAXMVAKLVSEVKOL"
complement(1452. .1745)
/gene="as13409"
complement(1452. .1745)
/gene="as13409"
/note="ORF_ID:as13409"
/codon_start=1
/transl_table=1
/product="nicotinamide nucleotide transhydrogenase,
subunit alpha"
/protein_id="BAB75108.1"
/db_xref="GI:17132542"
/translation="MTEALIALFVPLASPIGEIVINKTIPTLHPIMSGNSAIGSI
AVIGAILASGERNNLSYTLGLIANTLATVAVVAGFLVTRMLQMPKKKEVXA"
complement(1785. .3020)
/gene="a113410"
complement(1785. .3020)
/gene="a113410"
/note="ORF_ID:a113410"
/codon_start=1
/transl_table=1
/product="nicotinamide nucleotide transhydrogenase,
subunit alpha"
/protein_id="BAB75109.1"
/db_xref="GI:17132543"
/translation="MFLISKVADPAIRIYLSGSDENRIAAKEIEVCERRVALNPPT
VARLQKQLEWEEVAGAGERSFFDDAAVVAAGATVDSGKLWGETVLLKVSPPDR
EDGRLEIDLEEGVALISFLNPLDNPVAAQOLVARKVTALSMELIPRTTAQSMIDAS
SOASIAKGVALLIAAALPKYFPLMTAAAGTAPAKYFIMGAGVAGLOALATARSA
VLEANDIRPAVKEEVOSIGAKFEVVKLEBRTTAAGVAKESISASRORQEVAAEHX
NSDVITTRNQVGRKRAPVLYAEVNAQMKRGSVYVDLAAGCGNCACTDQCKITVNG
VTIIGLVPSEMPHASQLYSKNLTSLQLLKDKALQVDFDDIIDAACITHAGET
RNQRKDALQASTQVLVDR"
3627. .4199
/gene="a113411"
3627. .4199
/gene="a113411"
/note="ORF_ID:a113411
hypothetical protein"
/codon_start=1
/transl_table=1
/protein_id="BAB75110.1"
/db_xref="GI:17132544"
/translation="MLLYVNLVVISIKLPMMRRLTMRRLLSGLAVTSCVIAIGLPTTSA
QSLPFTLFSVCSEKSNQLPRLIDFGQGSNSVDRIYLRIPAKQKLAQAQALITPPHY
KGSFPTKQILIRVGDKKVPGEVMADEGRGLLEIFPEEPVADGSRVELVLSNVRNPAF
GSEFVFNKOISPDGVPLRLRIGTWIISIN"
4346. .4480
/gene="fp134"
4346. .4480
/gene="fp134"
/note="ORF_ID:as13412"
/codon_start=1
/transl_table=1
/product="50S ribosomal protein l34"
/protein_id="BAB75111.1"
/db_xref="GI:17132545"
/translation="MQRTLGSTNKKRRKRTSGFRARMTPPGRNVIRARRRGHRLSV
"
4506. .4928
/gene="a113413"
4506. .4928
CDS


```

/gene="a1r3413"
/note="ORF ID:a1r3413"
/codon_start=1
/translate=1
/product="ribonuclease P"
/protein_id="BAB75112.1"
/db_xref="GI:17132546"
/translation="MALPKANRLKSRHDPQAVPREGLRNNSHPTLRALKPSARKSS
LDTAAQDPVDEVONI PSTLIGVISITKUSKAIVNRI KROQTAAHQQLPRLAIGM
KLVIVKPTAKSKSCSQPLQELBOLAEVLKHS"
4915..5298
/gene="a1r3414"
4915..5298
/note="ORF ID:a1r3414"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75113.1"
/db_xref="GI:17132547"
/translation="MGIRESVYEGGPHIGDLINLNGITLVGIPLAAGATYRALML
KRIITRRISVMGMMGRSDVYSEITTKIVKVPGLGIWGDVVTLNKGSLERA
VENFREVDYDINKIAKNPQYTAK"
5513..6655
/gene="a1r3415"
5513..6655
/note="ORF ID:a1r3415"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75114.1"
/db_xref="GI:17132548"
/translation="MDFGIGFLSNNTMLPIIDFYSVBSYGLAIVALTLLIRFALYR
LSGQIRNMRKMIQVPLQKMAEI KERYKDPQKQOEEMNAVQKEFGNPLAGCPL
LIMPLVALFATLRGSPFAGVSVNLQIFPSEQIERIQPQAFAPRONIYADGSH
TKVTALIPGNKLAIVGKTKIYQYLEGKFPVLLAHEPSKLSPEMKITKEDRKI
DAEGNTEALQPDGNTVTCGTIGLAADKGPFLDALGRVADIDPNTIHDYMIYFP
GISLYSQMLSGNSSGSGNPQDTPNKKITFVIFSGFLFPPLPAGVLMKVNIGNVQT
LQTYILSRPLPEIQIKIVETOKVAVAKTLPFPKSKSKKTTG"
6659..7177
/gene="a1r3416"
6659..7177
/note="ORF ID:a1r3416"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75115.1"
/db_xref="GI:17132549"
/translation="KSNNPWQSGQOMLOSLELTGVSAETHSGSLFETASQSHNDSQEI
DGMWLTIDETNLNPOQIQTULIAGDSVLDAIQYLVNSTLINIQOEGQASVYVELNRYR
VKQAEIQIAETAEQVRSVSTGOEIEIKLSAERLVAHTFLKDFGDLFETFSRKEPH
RHVVPRVAVNEL"
7339..7848
/gene="a1r3417"
7339..7848
/note="ORF ID:a1r3417"
/hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75116.1"
/db_xref="GI:17132550"
/translation="MDAIFIPQLTKAPERTEEVQKFLPGLFETLTPVGLVRIQHG
NLYQVQAQBAIITCTCRCLQYORLAVNTQEVLMDEAVQEQNLPLERVAEMD
LLETVPNGYFVSEWLYEQMCLALPQRQCDINCIGIILDNSGESSBGGIDNRMSGL
KALKQLPG"
complement(7972..9447)
/gene="a113418"
complement(7972..9447)
/gene="a113418"
/note="ORF ID:a113418"

```

```

hypothetical protein"
/codon_start=1
/translate=1
/protein_id="BAB75117.1"
/db_xref="GI:17132551"
/translation="MTFMPFLLLGIATVLMVORSVANIITKPTIMLMLVMTPLVILT
GMTLVYGVKQPPPALIFWPSIICLLYLVLMOMGOLPDDQTAQOTTESGSAHPT
ABVPVRPIEPTBEQOLRBCPMSIYYONIEYRPAVISCROLPPTPOAXOORAN
TEQOFDRFLLLFOESFNKRPFFVLPNSOAKANKRSKSEQLTRPGLALLVATLYT
TTLVGAIKIAGIDPTRLQSDPKLFGSLPYALMLTILGHEGNYLTARPKIRSTLP
YFPIPFLEGTGAPIQMSPIPNRKALPDVGIAGLAFIATLPVIVGALNSDLVP
LTENTSILNPDALNPXYSIIVALLAKLIGLSALTAKLIDLHPVAVAGFLIVTALN

Alignment Scores:
Pred. No.: 1,176-67 Length: 333550
Score: 1606.00 Matches: 335
Percent Similarity: 62.204 Conservative: 101
Best Local Similarity: 47.794 Mismatches: 227
Query Match: 44.954 Indels: 38
DB: 1 Gaps: 9

US-10-089-514-2 (1-686) x AP003593 (1-333550)
QY 1 MetArgThrLeuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
      ::::::::::::::::::::
DB 33355 ATCAAAACATTATTTATTCACACTGACCTTATATCTTTATCTTACACGATGATT 33414
QY 21 GlyIuAlaThrGlyGlnProProValValProAsnAsp---AlaAspTPSerArg 39
      ::::::::::::::::::::
DB 33415 GCTGAAGTCACGAGAGAGTGCCTCTTATTCATATGACAAATTAAGATTGCTATGAA 33474
QY 40 LeuProValGluAspPheAspAlaIleValIleSerProGlyProGlySerProAspArg 59
      ::::::::::::::::::::
DB 33475 TTAACAACATTGTTTATATATATATATGATTCCTCGGGCGCTGCTGCAAAAC 33534
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAsp---SerGlyLeuProValLeu 78
      ::::::::::::::::::::
DB 33535 CCTAAAGATTTCGGTATTTGTCACAGATACATCAAAACCTATTATGATGTTCCCTTGCT 33594
QY 79 GlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThyValGlyLeuAla 98
      ::::::::::::::::::::
DB 33595 GGTGTTGCTTGCTGTCACAGGAATGTCATTATATATGCTGATGCTGCAACGCA 33654
QY 99 ProGluProMetHisGlyArgValSerGlyValArgHisThrGlyGluAspValPheArg 118
      ::::::::::::::::::::
DB 33655 CCAGAACCTTAACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 33714
QY 119 GlyLeuProSerProPheThrAlaValArgTYrHisSer---LeuAlaAlaThrAspLeu 137
      ::::::::::::::::::::
DB 33715 GAATTCCTTACACCTTTTCAAGTTGTCGCTATCATTTCTTGCTGCTGCGGAAGATTTG 33774
QY 138 ProAspGluLeuGlnProLeuAlaIleTTPSerAspArgGlyValValMetGlyLeuArgHis 157
      ::::::::::::::::::::
DB 33775 CCAGTGTGTTTGAAGAAAGTACCTGGAAGTGAAGAGGCTGTGATGAGGCTGCTCAT 33834
QY 158 ArgGluLeuProLeuTTrpGlyValGlnPheHisProGluSerTTrpGlySerAspPheGly 177
      ::::::::::::::::::::
DB 33835 CGTCATTTCGCAATGTCGGGGGGGCGAGTTTCCACCCGAAATCATATGCTGATGAGTAGCA 33894
QY 178 ArgGluIleMetAlaAsnPheArgSerLeuAlaLeuAlaHisHisArgValArgArgHis 197
      ::::::::::::::::::::
DB 33895 TATTAATTCCTGAAAATTCCAA-----GCCATTACGGAAGATTCACGCGCAAAAT 33948
QY 198 GlyAlaAspSerPro----- 202
      ::::::::::::::::::::
DB 33949 GGTAGACATCGCGCCAAATTAAGAAAGCAAAAGTTTCTATTTCTCTTAACACACTATT 34008
QY 203 -----TyrGluLeuHisValAlaArgValAspValLeuProAspAla 216
      ::::::::::::::::::::
DB 34009 CCAGCGCCATCCCAAGAAATTTACTGTTGCTACCAAAACATGATTTATACCTTGAATACA 34068
QY 217 GluGluValArgArgGlyCysLeuProGluGluGlyThrThrPheThrLeuAspSerSer 236
      ::::::::::::::::::::

```

Dh	34065	GAAACGACATTGTTCATCTTTTGGTGAAGATTCATGCTTTTGGCTTGATGATGAC	34128
Oy	237	SeValLeuenglIyAlaSerArgPheSerPheLeuGIyAAspAspArgIyProLeuAla	256
Dh	34129	CTTGAGAACCAAGGATTTATCCCGTTTCTCTTTTATGGAGAAAGCGGGGGGATTAACACT	34188
Oy	257	GIuTyLeuThrTyArgValAlaAlaSerGIyValValSerValArgIySerAspGIyThr	276
Dh	34189	TTACTCGTTTCTTATCTGACCCCAATCTCAAGAAATTCGATTCACAAATCG--GGTAAG	34245
Oy	277	ThrThrArgThrArgAAspProPhePheAsnTyLeuGluGluLeuGIuArgAArg	296
Dh	34246	GTGACAGCGCACTCAAGAGAGATTTTGTGACTATCTCAAGCGAAATTAATCGTCCCGT	34305
Oy	297	ValProValAlaProGluLeuProPheGluPheAsnLeuGIyTyArgIyTyLeuGIy	316
Dh	34306	TGCCCA--TCTGATGAGCTACCTTTGATTTAACGTGCGTTTGTGGGCTACTTGGT	34362
Oy	317	TyTGIuLeuLysAlaGIuThrThrGIyAspProAlaHisArgSerProHisProAspAla	336
Dh	34363	TATPAACTGAAAGGCAAAATCCGGGCTCGGGTTGAGAGATCCCTCATTAATTTCCCGATGCT	34422
Oy	337	AlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisArgIuGluGIyCysCysTyLeu	356
Dh	34423	ATCTTGATTTTAAACGATTCGCTCGATGTGCTTTTGAACCAACAAAGAAACCACTTA	34482
Oy	357	LeuAlaLeuAspArgAArgIyHisAspAspGIyAlaArgAlaTyPheLeuArgIuThrAla	376
Dh	34483	GTTTAACTCGCCCGCTGGGGTGAACAGCAACGACCCAAACCTGGTTTGAACTAAGTGA	34542
Oy	377	GIuThrLeuThrGIyLeuAlaValAlaArgAlaProAlaGIuProThrProAlaMetValPhe	396
Dh	34543	AAACGCTGCGAGCACTC-----CCGCGTCTTCTCCCTCCCTCGTCTCT	34584
Oy	397	GIyLeuProGluAlaAlaAlaGIyPheGIyProLeuAlaArgAlaArgHisAspIyAsp	416
Dh	34585	AGTACTACATCTACCCCGTTACTTTC-----CGTTAAAGTCGGTCATACAG	34632
Oy	417	AlaTyTyLeuLysArgIleAspGIyCysLeuLysGIuIleArgAsnIyGluSerTyArgIu	436
Dh	34633	ACATTTAAAGATGATATTCTTCAAGTGTCTCAAGAAATTCGTGAGGGGAAACTTACCA	34692
Oy	437	IleCysLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTySer	456
Dh	34693	GTTTGCTGCTATCATTTATGACGACAGACAACTTGATCATTTGAATTTTATCGC	34752
Oy	457	AlaLeuArgAlaIleSerProValProTyGIyAlaLeuLeuGIuPheProGIuLeuSer	476
Dh	34753	CGATTTCAGCAAGTTAATCCGGCTCCCTACTCGCTTTTCTACGGTTTGGTAGAGGGG	34812
Oy	477	ValLeuSerAlaSerProGIuArgPheLeuThrIleGIyAlaAspGIyGlyValGluSer	496
Dh	34813	ATGCCCGTCTTCTCCCGAGCGCATTTCTGCAATTTGACCGGTCAAGCTGGGTGAAAC	34872
Oy	497	LysProIleLysGIyThrArgProArgGIyIyThrAlaGIuGluAspGIuArgLeuArg	516
Dh	34873	AAGCCGATCAAGGAACTGTACCGCGGGAGAGACTCCGAAAGAAATTTTATCTCCGG	34932
Oy	517	AlaAspLeuAlaGIyArgGIyLysAspArgAlaGIuAsnLeuMetIleValAspLeuVal	536
Dh	34933	GAAAGTTTGCGTACAGCGAAAGAAAGACCGATCCGAAAACTTAATGATTGTGATTTATTA	34992
Oy	537	ArgAsnAspLeuAsnSerValCysAlaIleGIySerValHisValProArgLeuPheGlu	556
Dh	34993	CGCAACGATTTTAGGCGCAGTTTCCCAAGTCGGAGATTCATGTCTCCCAATTAATGAT	35052
Oy	557	ValGIuThrTyArgAlaProValHisGluLeuValSerThrIleArgGIyArgLeuArgPro	576
Dh	35053	GTAAAGAAATTATGCAAGCGTGATCAACTTGTCAAGCATATCCGTGGTTATTATGCC	35112
Oy	577	GIyThrSerThrAlaAlaCysValArgAlaAlaPheProGIyGIySerMetThrGIyAla	596
Dh	35113	GACATGACGCAACCGATCATTTCTGCTGGCAATTTCCGGGCGCGTCAATGACAGGTGG	35172

[illegible]

Qy	641	AlaaspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAaspGlnGlu	660	/organism="Streptomyces avermiltillis MA-4680"
				/mol_type="genomic DNA"
Db	13689	ACCGGTACGGGACACGGTCATCGGACCGGGGGCCCATCTCTCCGACTCGAACCCGGAC	13748	/strain="MA-4680"
				/db_xref="taxon:227882"
Qy	661	GluGluPheThrGlnThrValValValAlaArgAlaMetValThrAla-----Leu	677	/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinus"
				208..921
Db	13749	GAGGAGTTTCAGCAGGTGCTCTCAAGCGCACGCCGCTGGTGGCGGCGCACGGGCTGCTG	13808	/gene="orf-1"
				208..921
Qy	678	AaspGlySerAlaValAla	683	/gene="orf-1"
				208..921
Db	13809	GAGGGGGCCGCGGAGGCC	13826	/note="SAV944 not involved in biosynthesis of avermectin"
RESULT 14				
AP005025/C				
LOCUS	AP005025	298300 bp	DNA	linear
DEFINITION	Streptomyces avermiltillis genomic DNA, complete genome, section 5/30.			
ACCESSION	AP005025	BA000030		
VERSION	AP005025.1	GI:29604584		
KEYWORDS	Streptomyces avermiltillis MA-4680			
SOURCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.			
ORGANISM	1			
REFERENCE	1			
AUTHORS	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kikuchi, H., Shiba, T., Sakaki, Y. and Hattori, M.			
TITLE	Genome sequence of an industrial microorganism Streptomyces avermiltillis: deducing the ability of producing secondary metabolites			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	98 (21)	12215-12220	(2001)
MEDLINE	21477403			
PUBMED	11572948			
REFERENCE	2			
AUTHORS	Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. and Omura, S.			
TITLE	Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermiltillis			
JOURNAL	Nat. Biotechnol.	21 (5)	526-531	(2003)
MEDLINE	22608306			
PUBMED	12692562			
REFERENCE	3			
AUTHORS	(bases 1 to 298300)			
TITLE	Omura, S., Ikeda, H., Ishikawa, J., Hanamoto, A., Takahashi, C., Shinose, M., Takahashi, Y., Horikawa, H., Nakazawa, H., Osonoe, T., Kushida, N., Director-General of Biotechnology Center, Shiba, T., Sakaki, Y. and Hattori, M.			
JOURNAL	Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-ku, Tokyo 151-0066, Japan			
COMMENT	(E-mail: bio@nitech.go.jp, URL: http://www.bio.nitech.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)			
	This work was done in collaboration with Haruo Ikeda (*1), Jun Ishikawa (*2), Akiharu Hanamoto (*3), Chigusa Takahashi (*3), Jun Shinose (*3), Hiroshi Horikawa (*4), Hideaki Nakazawa (*4), Tomomi Osonoe (*4), Norihiro Kushida (*4), Hisashi Kikuchi (*4), Tadayoshi Shiba (*5), Yoshiyuki Sakaki (*6, *7), Masahira Hattori (*1, *7) and Satoru Omura (*1, *3).			
	Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.			
	*1 Kitasato Institute for Life Sciences, Kitasato University			
	*2 National Institute of Infectious Diseases			
	*3 The Kitasato Institute			
	*4 National Institute of Technology and Evaluation			
	*5 School of Science, Kitasato University			
	*6 Institute of Medical Science, University of Tokyo			
	*7 RIKEN, Genomic Sciences Center			
	Following url is also available.			
	http://avermiltillis.lis.kitasato-u.ac.jp.			
FEATURES	Location/Qualifiers			
source	1..298300			

Db 267013 GGCACGGCGTTCGACTACCTGAAACGGCAGTGTGTGACCCGTCAGTGGAT---GCCAC 266957
 Qy 301 oGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuVal 321
 Db 266956 GGAATCCCGCTTCACCTTACCGCGGAGTATGTGGCTACTTCGCGTAACGAGGTAAGGC 266897
 Qy 321 aglUtrThrGlyAspProAlaHisArgSerProHisSproAlaAlaPheLeuPheAl 341
 Db 266896 CGACTGGGTTCCCTCAACCGTATCCGGCCCGAACTCCGACCCCTGCTGCTTCCG 266837
 Qy 341 aAspArgAlaAlaLeuAspHisGlnGlyGlyCysCysTyrLeuLeuAlaLeu---- 359
 Db 266836 CGACCGGCTGATCCCATGTGACATCAGACGAGACCACTACCGCTGCTGCTCGCGCA 266777
 Qy 360 ----AspArgArgGlyHisAspArgGlyAlaArgAlaTyrLeuArgGluThrAlaGluTh 378
 Db 266776 GAACACCGGAGGGGCGCACAGCAAGCGGCC--GACTGGCTCAACCGCGGATGCGCA 266720
 Qy 378 rLeuThrGlyLeuAlaValArgAlaPro---AlaGluProThrProAlaMetValPheG 397
 Db 266719 GCTCAGTTTGTGTCTCCGCGCAAGCCGACCGCCCGCGCGCCGACG----- 266671
 Qy 397 ylleProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspYAspAl 417
 Db 266670 -GCAACCCCGCTCGATGCCGTGAGCGCGTGGCTGCTCCG-----GACCGTACGAC 266621
 Qy 417 aTyTyrLeuValArgGlyLeuAspGlyCysLeuGlyGluTyrAlaArgAsnGlyGluSerTy 437
 Db 266620 CTACCTCGCGGACACTCGGACCTCGCAAGCGGAACTCAAGACCGGACCGATTACGAGAT 266561
 Qy 437 eCysLeuThrAsmMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAl 457
 Db 266560 CTGTCTCAACCAACCGCGCCAGCGCTGCCCGCCCTACAGCCCTTACGATCCGGGT 266501
 Qy 457 aLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerVa 477
 Db 266500 ACTGCGTCTGTCGACCGCGCGCGCTTACCGCGCTTACCGGTTCGTAACCTCGATGT 266441
 Qy 477 lleuSerIleSerProGluValPheLeuThrIleGlyAlaAspGlyGlyValGluSerTy 497
 Db 266440 GGCCTGCTCTCCCGCGAGGCTTTCTCGCATCACCGCGACGCGCTCCCGCGAGCGAC 266381
 Qy 497 sProIleGlyGlyThrArgProArgGlyGlyThrAlaGluGluAspGluValArgAl 517
 Db 266380 ACCGCTCAAGGGACCGCGCGCGCGCGCGAGCGCGCGAGAGAGACCGCGCTACCGGA 266321
 Qy 517 aAspLeuAlaGlyArgGluValAspArgAlaGluAsnLeuMetIleValAspLeuValAr 537
 Db 266320 CGGCTCACCAACGAGACAAACCGCGCGGAGAACCTGTATGTTCACCTGCTCCG 266261
 Qy 537 gaAspLeuAsnSerValCysAlaAlaIleGlySerValHisValProArgLeuPheGluVa 557
 Db 266260 CAACGACCTGGCGCGGCTGCGCGACCGGAGCGGTGAAGTCAACCGCTGATGGCGAC 266201
 Qy 557 lglUtrTyrAlaProValHisGlnLeuValSerThrIleArgGlyValArgLeuArgProGl 577
 Db 266200 CGAGACGTGCGCACCGGTGACCAAGTGTGCTCCACCGTGAAGGGCGGCTGCGCGAGGG 266141
 Qy 577 yThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPr 597
 Db 266140 CATCGGAGCGGTGAGCTGGCTTACGTCTTCCCGGAGTTCGGTGAACCGGGGCGCC 266081
 Qy 597 oLybLybArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSe 617
 Db 266080 GAAACCTGCGCACATGATGATCATTCGCTGCGAAGACGAGGCCCGCGGGGTGTACTC 266021
 Qy 617 rglValAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgTh 637
 Db 266020 CGAGGCGCATGGGCTTCCGGGTGCACTGGAGGGCGCGGACCTCGCCACGCGCAC 265961
 Qy 637 rIleValIleuAlaAspGlyGlnAlaGluPheGlyValGlyValAlleValSerLeuSe 657
 Db 265960 GGCGGTGTTACCGACGCGGAGATGCACCTCGGCGCGGGGCTGCGATCTCTCGGCTC 265901

Qy 657 rAspGlnGluGluPheThrGluThrValValIleValArgAlaMetValThrAlaLe 677
 Db 265900 CGATCCCGTGGCGGAGTACACAGATGCTGTGAAGCGCGCACCGCATGGCGCC-- 265843
 Qy 677 uAspGlySerAlaValAlaGlyAlaArg 686
 Db 265842 -CACCGGACCGGATGCGCGCTCGCG 265816
 RESULT 15
 AP005217
 LOCUS
 DEFINITION
 Corynebacterium efficiens YS-314 DNA, complete genome, section 4/11
 VERSION
 AP005217 BA000035
 AP005217.1 GI:23492722
 KEYWORDS
 ORGANISM
 Corynebacterium efficiens YS-314
 Corynebacterium efficiens YS-314
 Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 Corynebacterineae; Corynebacteriaceae; Corynebacterium.
 REFERENCE
 AUTHORS
 1 Nishio, Y., Nakamura, Y., Kawarabayashi, Y., Usuda, Y., Kimura, B., Sugimoto, S., Matsui, K., Yamagishi, A., Kikuchi, H., Ikeo, K. and Gojobori, T.
 TITLE
 Comparative complete genome sequence analysis of the amino acid replacements responsible for the thermostability of Corynebacterium efficiens
 JOURNAL
 MEDLINE
 PUBMED
 22723752
 12840036
 2 (bases 1 to 300750)
 REFERENCE
 AUTHORS
 Kawarabayashi, Y., Yamazaki, J., Hino, Y., Kikuchi, H. and Director-General of Biotechnology Center.
 TITLE
 JOURNAL
 Submitted (17-MAY-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10, Shibuya-ku, Tokyo 151-0066, Japan (E-mail: bio@nitech.go.jp, Tel: 81-3-3481-1933, Fax: 81-3-3481-8424)
 COMMENT
 Kawarabayashi, Y. is officially affiliated with the National Institute of Advanced Industrial Science and Technology, Tsukuba, Ibaraki, 305-8566 Japan
 Nakamura, Y., Ikeo, K., Suzuki, M. and Mashima, J. are at the National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan
 Itoh, T. is at the Japan Biological Information Research Center, Koto-ku, Tokyo, 135-0064 Japan
 Yamagishi, A. is at Tokyo University of Pharmacy and Life Science, Nishio, Tokyo, 192-0392 Japan
 Nishio, Y., Usuda, Y. and Sugimoto, S. are at the Ajinomoto Co., Inc., Kawasaki, Kanagawa, 210-8681 Japan
 The other authors are at the National Institute of Technology and Evaluation, Shibuya-ku, Tokyo, 151-0066 Japan.
 Location/Qualifiers
 1..300750
 /organism="Corynebacterium efficiens YS-314"
 /mol_type="genomic DNA"
 /strain="YS-314"
 /db_xref="taxon:196164"
 252..620
 /note="CB0885"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAC17695.1"
 /db_xref="GI:23492723"
 /translation="MDQAPSGGLDNCVGYIFPDTPVPMILFPVGTSGNRKIVLDFN HRFNANKPRILITPLMSGGCLFGAVIGLFTWHVHVVCVFCBFPDGSRTLAAC LTHASRTMKPCIQGGIVANG"
 529..1936
 /gene="rna_16S"
 529..1936
 /gene="rna_16S"
 /product="16S ribosomal RNA"

CDS
/note="CE-RNA01"
/complement (2042..2215)
/note="CE0886"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17696.1"
/db_xref="GI:23492724"
/translation="MSSQTPNSMPRTYHHHLDGVADPWMMIHPIITTCWQHPRGSNT
NNPHEHSPMWVY"
2399..5481
/gene="trna_23S"
/gene="trna_23S"
2399..5481
/gene="trna_23S"
/product="23S ribosomal RNA"
/note="CE-RNA02"
5606..5725
/gene="trna_5S"
/gene="trna_5S"
5606..5725
/product="5S ribosomal RNA"
/note="CE-RNA03"
5925..7088
/note="CE0889, similar to AK063829-1|CAC25156.1| percent
identity: 79 in 387 aa"
/codon_start=1
/transl_table=11
/product="putative aminotransferase"
/protein_id="BAC17699.1"
/db_xref="GI:23492727"
/translation="MRNKYVERLRPFGETTIFATWTORANEAGAINLGGCFPDEGPA
PMLAISQQLIGNNQYSGRGDLGRTAVASRRHAKFKLDYDPDTEVLIYVGTATAI
SASVLIGVEGDEVILEPYDAYAAIALAGATRYAVPLKEEDNTWVLDGTAVAV
TKTSMIVNSPHNPITGSVSKSLALAGIARAYDLVLSDEVHEHLPFDKXSHAV
ASIPGMMDRTITYSSAKTPNTYGMKTGALAPRILEAVIRAKQMSYVGATPRPRA
VAAVAHKKRPVQKMRKGLONKSDIITDLITQGLVTHSHGTYFVAADIGEDGAEF
CFDLIRVGVAAIPVOAFVDPKPEQSSKVFACQKQBDTLREAAQRLRAAGSL"
complement (7180..7821)
/note="CE0890"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17700.1"
/db_xref="GI:23492728"
/translation="MTHSFSDVDEAVARKNNEFLRDTKRLQISAPFPGHVLIAIGVL
VYLDAVGWGMWVLIWGIAMALSFVMIPLPRQVGSASISLVNTYVPLVAVVAENPR
DVVLLAVNTNVDETIIPRKGLAVRTINRLGVDRKLGERSVAVNSGRRTTRDQHM
DEISMPITWGTIDADVRRKERTIPHDILNNKLEKRNRLDEVKTRMNLVL"
complement (7822..9495)
/note="CE0891, similar to AL583924-93|CAC31112.1| percent
identity: 75 in 547 aa"
/codon_start=1
/transl_table=11
/product="putative DNA helicase"
/protein_id="BAC17701.1"
/db_xref="GI:23492729"
/translation="MAFGDGLIVOSDKTVLLETGDLAGEARALAPFALERAEH
VHTYRITPLALNNARAAGDAEQVNDVLRYSFPPOPILVNAETMGRYGVNRMK
HPNAGLILBSESPAILIVEISRHKIRPMIGQOUDPTIAVHPSERGLKQELKIKMP
AEDLAGVDEAHPIALSTEVEDWSLRDYOQVADSFWEQSGSVVLPFGAGTWTGA
ASMARAOATLLIVNTVAGRWKDLRLRTTLTBEIGYSGERKEIRPVATATVO
VTRRTKGEYKALEIPSRDWGLIYDEVHLLPAPVPMTSDLOSRRRLGITATVED
GREGVDSLIGPRYDAPMKDLEAGFIATADCEVETMTDAERMVYVTAETADNR
LATATKTRVVRRIIEOHGQPTLIIGAYDQLELGAEPDAIYDQCTSNKREEL
PFRFSGELNVLVSKYANFSIDLPRANAIVOSGTGFSROEBAQRLGRLPKADGG
EAHFYVSVDITLDTFYAAHRORFLABQGYARIIDADDLLFPITEKET"
complement (9548..11851)
/note="CE0892, similar to AE006976-11|AAK45126.1| percent
identity: 36 in 741 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17702.1"

CDS
/db_xref="GI:23492730"
/translation="MPCFPFSQIRQTKVTSAPKLGDIIMKQDSPVPTLTGWLQOLD
OLSGILRRNPDPALPLPGLGSLARLQASIRIOAVHKLSTALELVAEANAAGEL
NPITAPVEVERLRHRLRRBSAAYPTAEQIGELNRLROLALVGRGRFTILPEMTALP
TNWQILPREHSGSLTFGEASQSIEALPARHKITVTLVTSNGLGTRDAALSADPARP
ITQLIKAGLSRVDEQIVKLPGVRRVVEGQEDTAMRVVPEBSAVPGADGIGLAGL
EVARLQRMILIDALSTAPASTKEGALGVMMVTRLREIQVDQVAVRVSLCVSGSL
RGVDDPLPADDDGDYIAPTVADEMQLQATLAEQALHMRGMTOYVAAVLVEADE
KKKPIHLSPASRDALPETRAMILOSILTRVEEDLEADLPHHPLAASRIPRETIRH
LLEBARWGAVAGGVTSAPVRLTOAPRGVIEIDIVAVAPRPVDYFIVQADYTV
WYGPQLQPMQKTIQGLADBSRGASVYRISBASLRHAMDGLTARETEDTQSHA
TGLPOSVALTLODIARRGHTLRGSPALCYRSDDPALHSAVEALGALRDIATPVA
VSNASHLIEVITLRRAGQVPAEDALGASLDLAPRPAPVPAEQPRSGAPLDESRI
AAVAIRREBEVARTGVSDQPTLAVIQAAVRCGRVTTLGFEVDKQGVAVHRTVPLTVNA
GVQVLDSESTGAVHRFMLRITEVIVD"
11836..12036
/note="CE0893, similar to AL583924-91|CAC31110.1| percent
identity: 50 in 62 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17703.1"
/db_xref="GI:23492731"
/translation="MGNNANVEKKGYVDPAMPDHPDSEGHVTEILAPYAGASSPWGD
DNEFPVPATTIGYVHPYTRINR"
complement (12229..12822)
/note="CE0894, similar to AX133781-1|CAC38957.1| percent
identity: 71 in 195 aa"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="BAC17704.1"
/db_xref="GI:23492732"
/translation="MLFMGRSHAKRTAPFVKMAATTVAPFGAAVAVAPAPASAAPSDW
DLIANCEGGMNAINTGNGYHGLDPSSTTRWAKYGDDEAPAPAYATREQQAVABRT
LAGQGMAMPACSSKLGNSAPTPNNLNPAPAPAPAPAPARATVDTNTNPVVGSS
DINTQGIYDAVSGTLAOGVQVPAEIQAHVVAFTGR"
13342..13725
/note="CE0895, similar to AE006977-4|AAK45135.1| percent
identity: 52 in 127 aa"
/codon_start=1
/transl_table=11
/product="putative cold shock protein"
/protein_id="BAC17705.1"
/db_xref="GI:23492733"
/translation="MPGVTKVYDPERGGRFVSNQGEDCFYKQVLPKCYTELHOG
RIEFPFAPARKGPQLARVAVLETPRRRPQHTYKPEELNGIISDMVTMLEGTQPALIK
GHYFPRKVGEOYAKILIRAVAKYLEA"
complement (13881..14408)
/note="CE0896"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAC17706.1"
/db_xref="GI:23492734"
/translation="MASRKYKKNFIQATLLVAVVVVVISVVFQNNMNNRPEPE
TVSIYASSPAGIEVFPYSICRGVCAENEPVTLVEGDEELRLTIPIIHDHMYL
LTIYDPAANDRFYTSYETTEVYVPGSDVPTBEGGERRLVVEVTSWICITDNGE
ETPYTTTSLATVEQ"
complement (14424..15254)
/note="CE0897, similar to AE001874-1|AAF09704.1| percent
identity: 40 in 267 aa"
/codon_start=1
/transl_table=11
/product="putative glutamine cyclotransferase"

Alignment Scores:

Pred. No.:	Length:	300750
Score:	1541.00	Matches: 342
Percent Similarity:	60.71%	Conservative: 63
Best Local Similarity:	50.52%	Mismatches: 206
Query Match:	43.13%	Indels: 60
DB:	1	Gaps: 13

US-10-089-514-2 (1-686) x AP005217 (1-300750)

QY 1 MetArgThrIleuLeuLleAspSerPheThrHisValLeuPheGlnTyrIle 20
184888 GTGGCGCTACTGATCATCGCAATATGATTTCTTCCCTTCAACCTTCGCGCATATGTC 184947
DB 21 GlyIuAlaThrGlyGlnProProValValProAspAlaAspTyrPseArgLeu 40
184948 GAGAGGTCACCGGTGGGACGACGAGTGGTCCCATATGCG-----CAATC 184998
QY 41 ProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArgGlu 60
184999 GACGAGACGCTGTTGACGCGCATCATCTCACCGCGCCCGCCACCCCGGGGTGGC 185058
DB 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGlyVal 80
185059 GCGGATTTCCGCGAGCTCCCGGGGTGATCGAACGTGTGTCGTCGCGGTCTTCGGGGTGC 185118
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
185119 TGCTGGGACATCAGGAGCATCGCTCGCCACGCGCGCGGTGGCGCGCGCTC 185178
DB 101 ProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
185179 CCGGTGCACGCTGACGTCACCGCATCCACACGAGGATCCGAGCTTTTCGACGCGCATC 185238
QY 121 ProSerProPheThrAlaValArgTyrHisSerIleuAlaIleThrAspLeuProAspGlu 140
185239 CCGCCCGAGTTCCAGCGCGCTTACCTCCCTGGTGGCCACCGCATCTCGCGCGCGGA 185298
DB 141 LeuGluProLeuAlaTyrPseAspAspArgValValMetGlyLeuAlaGlnGlyLeu 160
185299 CTGAGAGTCACCGCGAGACCGGGAGCGGTATGCTGCGCTGCGCCACCGCGAGCTG 185358
QY 161 ProLeuTyrGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIle 180
185359 CCCAGTGGGGGTGTGCGATTCACCGCAATCCATCGCGCGGAGTTCCGGGACCGGATC 185418
DB 181 MetAlaAspPheArgPheLeuAlaLeuAlaHisValArgAlaArgArgHisGlyValAsp 200
185419 ATGGCGAATTTTC-----CTGAGCTTGGCGCTGCGACGACGCG----- 185460
QY 201 SerProTyrGlnLeuHisValArgArgValAspValLeuProAspAlaGlnGluValArg 220
185461 -----TGGAGATCACCGAGCATGTGGTGAACAAGCCTGACCGCGCGCGGTGTC 185514
DB 221 ArgGlyCysLeuProGlyGlnGlyThrThrPheTyrLeuAspSerSerValLeuGlu 240
185515 GAGACGCTTCCCGCGGGTGGAGACGCGCTTCTGGCTCGATGATCCG-----CAG 185565
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyTyrLeuThr 260
185566 GGCACG-----ACCTATATGGGTGATCGCTCCGGCGCGCATGACGAGATCCGAC 185616
DB 261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280
185617 CACCGGGTGGGAGGG----- 185634
QY 281 ArgArgProPhePheAspTyrLeuGlnGluGlnLeuArgArgValProValAla 300
185635 -----GAGCTGTTGCACTGGCTGGATGATCTGGCTGCAACGCG-----GTGGCC 185682
DB 301 ProGluLeuProPheGlnPheAspLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320
185683 CCGGGGGTGTG-----GGTTTCCGCTGGGGTGGGTGAATCCTCGGATATGAGATGAG 185736
QY 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
185737 GCGGAATGCGGGTGAACAATCGCAGCTCATCGATCCGATGCCAGTGCACCTGATCTTC 185796
DB 341 AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysSerTyrLeuLeuAlaLeuAsp 360
185797 GCGGACCGGGCCATCGCATC-----GAACTCCGGCGGGGTGNGGCTCATGCGGCTCGCT 185850

QY 361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrPheAspGluThrAlaGluThrLeuThr 380
185851 GACGAGGGGAG-----TGGTTCGCGAGATGACCGCGGCTCGGGG 185892
DB 381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
185893 CAGCTG-----CGCCA-----CCCGCT 185910
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys 420
185911 GCGGCTGCGCGCGCGCGCGCGCGCGCTCACCGCTCGGGAGTACCGGACAGCTACCTGGAC 185970
DB 421 ArgIleAspGlyCysLeuLysGluIleArgAspGlyGlySerTyrGluIleCysLeuThr 440
185971 ATGATCGCGCGCGCGCGAGATTGATCACCGCGGGAGATCTCATAGATCTGCTGAC 186030
QY 441 AsnMetValThrAlaProThrGlyAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
186031 ACCGAGCTGCGCGGAGGTGAGGTGATCCCTGCGCGCTATGTGGCTGCGGCG 186090
DB 461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValIleSerAla 480
186091 GCCAACCCACCTCTTATGATCATCTCTGCACTGGGGAGATGGCGGTGCTGAGCTCG 186150
QY 481 SerProGluAspPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLys 500
186151 TCGCCGGAAGGTTTCATCAGCATCGAGGCTCGGAGGTTGGAATCCAAACCATCAAG 186210
DB 501 GlyThrArgProArgGlyGlyThrAlaGlnGluAspGluArgLeuArgAlaAspLeuAla 520
186211 GGCACCGCGCGCGGGGAGTACGAGCAGAGGAAGCGGCTTGTATCCGATCTTCACC 186270
QY 521 GlyArgGluLysAspArgAlaGluAspLeuMetIleValAspLeuValArgAspAspLeu 540
186271 GACAAACCCAAAGACCGCGCGCAAGACCTCATGATGTGACTGTCCGCAATGATCTTC 186330
DB 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
186331 GCCCGGGGGCGCAACCGGACAGGTGAGTGAAGAACCTTTTCAGCTCGAGACCTAC 186390
QY 561 AlaProValHisGlnLeuValSerThrIleArgGlyArgLeuAspArgProGlyThrSerThr 580
186391 GCCACGCTCCACAGCTCTGTAGCACCATTAACCGCCACGCTGAG-----GTAAAGACCCC 186447
DB 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600
186448 ATTGACTGCGTGGCGGCGCGCTTCCCGGTGTTCCATACCGGTCCACCGAAGATCCG 186507
QY 601 ThrMetGluIleIleAspArgLeuGlnGlyProArgGlyValTyrSerGlyAlaLeu 620
186508 ACCATGAGATCATGACAGAGTGAAGACCGGCGCGCGGTGATTCATTCGCGTGGCCGTG 186567
DB 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
186568 GGCTATTTCTCCCTTGATGGGCGGGGTGATCTGTGCATGTGATCCGACCGGTGGCTTAC 186627
QY 641 AlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGlu 660
186628 ACCCGCGGCTGCTGAGTACGGAGTGGGGGAGATCTGAGCTGCTGCGACCGCGCC 186687
DB 661 GlnGluPheThrGluThrValValIlyAspAlaArgAlaMetValThrAlaLeu 677
186688 GCAGATGGAGAGATCGGGTGAATGAGAGCCCTGCTGGGTCTGTC 186738

Search completed: October 5, 2005, 15:27:53
Job time : 10957.6 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 4, 2005, 21:38:23 ; Search time 1204.67 Seconds
(without alignments)
3371.006 Million cell updates/sec

Title: US-10-089-514-2
Perfect score: 3573
Sequence: 1 MRTLIDNYSFTNHLFOYI.....VKKAMVTLDDSAVAGAR 686

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=x1h
-Q=/cgr2_1/US10089514/runat_04102005_105743_8027/app.query.fasta_1.1621
-DB=N_Geneseq -QFMT=fastcap -SUFIX=ring -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=bloum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NOR=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USR=MSI0089514 @CCN 1.1_860 @runat_04102005_105743_8027 -NCPU=6 -ICPU=3
-NO_WARP -LARGEIOBERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARP_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: Geneseq_16Dec04:*
2: Geneseq_19808:*
3: Geneseq_19908:*
4: Geneseq_20008:*
5: Geneseq_20018:*
6: Geneseq_20028:*
7: Geneseq_20028:*
8: Geneseq_20038:*
9: Geneseq_20038:*
10: Geneseq_20038:*
11: Geneseq_20038:*
12: Geneseq_20048:*
13: Geneseq_20048:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3573	100.0	2061	AAAF6409	AAAF6409 4-amino-4
2	3573	100.0	2061	AAAF6409	AAAF6409 4-amino-4
3	3127.5	87.5	3305	ABZ69799	ABZ69799 S. venezue
4	3127.5	87.5	12391	ABZ69798	ABZ69798 Plasmid p
5	1487	41.6	1860	AAH6076	AAH6076 C glutami

6	1487	41.6	1983	4	AAAF72103	AAAF72103 Coryneb
7	1487	41.6	1983	4	AAAF71982	AAAF71982 Coryneb
8	1487	41.6	2005	10	ADN13458	ADN13458 C. glutam
9	1487	41.6	349980	5	AAH68527	AAH68527 C. glutami
10	1475.5	41.3	110000	10	ACF67367_49	ACF67367_49
11	1475.5	41.3	110000	10	ACF65387_1	ACF65387_1
12	1400.5	39.2	2019	10	ACF71773	ACF71773 Phototrab
13	1350	37.8	2046	9	ADBF6077	ADBF6077 Allostoc
14	1350	37.8	110000	9	ADB12064_00	ADB12064 Allostoc
15	1288	36.0	2760	10	ADB95077	ADB95077 A. thalia
16	1244	34.8	2151	12	ADN36903	ADN36903 X. albii
17	1244	34.8	55839	12	ADN36893	ADN36893 X. albii
18	1228	34.4	26309	4	AAAS9535	AAAS9535 Propionib
19	1228	34.4	26309	4	ACF64464	ACF64464 Propionib
20	1096.5	30.7	4496	2	AAAF8553	AAAF8553 Streptomy
21	1068.5	29.9	2523	8	ACAF2955	ACAF2955 Streptomy
22	1001	28.0	110000	6	ACA29107	ACA29107 Streptomy
23	859.5	24.1	110000	6	ABA90521_13	ABA90521_13
24	834	23.3	110000	6	ABO67196_3	ABO67196_3
25	834	23.3	110000	6	ABO69245_28	ABO69245_28
26	834	23.3	110000	6	ABO69245_29	ABO69245_29
27	795.5	22.3	110000	6	ABA03041_28	ABA03041_28
28	725.5	20.3	110000	6	AAAF5268	AAAF5268 Streptomy
29	657	18.4	2648	4	AAAF5683	AAAF5683 Streptomy
30	657	18.4	2648	4	ACF64612	ACF64612 Streptomy
31	623	17.4	1362	8	ACA42265	ACA42265 Streptomy
32	615	17.2	1431	11	ABD15129	ABD15129 Pseudom
33	579.5	16.2	1377	8	ACA53211	ACA53211 Streptomy
34	578.5	16.2	1101	11	ABD14809	ABD14809 Pseudom
35	570.5	16.0	110000	4	AAAF9682_17	AAAF9682_17
36	570.5	16.0	110000	4	AAAF9682_18	AAAF9682_18
37	570.5	16.0	110000	4	AAAF9683_17	AAAF9683_17
38	568	15.9	1377	8	ACA53954	ACA53954 Streptomy
39	566.5	15.9	1338	8	ACA44234	ACA44234 Streptomy
40	563	15.8	1333	8	ACA31740	ACA31740 Streptomy
41	563	15.8	1506	11	ABD02207	ABD02207 Pseudom
42	561.5	15.7	1365	8	ACA51013	ACA51013 Streptomy
43	561	15.7	1440	9	ACA21964	ACA21964 Streptomy
44	558	15.6	1503	8	ADA30238	ADA30238 DNA encod
45	557.5	15.6	1365	8	ACA51642	ACA51642 Streptomy

ALIGNMENTS

RESULT 1	AAAF6409	standard; DNA; 2061 BP.
ID	AAAF6409	
XX	AAAF6409;	
AC	AAAF6409;	
XX	AAAF6409;	
DT	22-JUN-2001	(first entry)
XX	AAAF6409;	
DE	4-amino-4-deoxychorismic acid synthase, papA, coding sequence.	
XX	Metabolite: benzene; chorismic acid; p-aminophenylpyruvic acid;	
KM	4-amino-4-deoxychorismic acid synthase; enzyme; papA; ds.	
XX	Streptomyces venezuelae.	
OS	Streptomyces venezuelae.	
XX	Streptomyces venezuelae.	
FT	Key	Location/Qualifiers
FT	CDS	1..2061
FT		/tag= a
FT		/product= "4-amino-4-deoxychorismic acid synthase"
XX	MO200123542-A1.	
XX	05-APR-2001.	
PD	05-APR-2001.	
XX	05-APR-2001.	
PF	29-SEP-2000; 2000MO-UP006783.	
XX	29-SEP-1999; 99UP-00276314.	
PR	29-SEP-1999; 99UP-00276314.	
XX	(MEIJ) MEIJI SEIKA KAISHA LTD.	
PA	(MEIJ) MEIJI SEIKA KAISHA LTD.	

XX		Yanai K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Midoh N,
PI		Murakami T;
XX		WPI; 2001-290517/30.
DR		P-PsDB; AAB82071.
XX		
PT		Transformant producing secondary metabolite modified with functional
PT		group e.g. benzene with nitrogen-containing substituent at para-position,
PT		PF1022, with ease at low cost, for application in pharmaceuticals and
PT		agrochemicals.
PS		
PS		Claim 11; Page 54-59; 83pp; Japanese.
XX		
CC		The present invention relates to a transformant having been modified so
CC		as to produce a secondary metabolite. The secondary metabolite has a
CC		benzene ring skeleton free from substitution at the para-position by a
CC		nitrogen-containing functional group, thereby enabling the production of
CC		a secondary metabolite with a benzene ring skeleton substituted at the
CC		para-position by a nitrogen-containing group. The transformant organism
CC		of the present invention has been produced by transferring a gene
CC		participating in the biosynthesis pathway from chorismic acid into p-
CC		4-amino-4-deoxychorismic acid. The present sequence is the coding sequence for
CC		4-amino-4-deoxychorismic acid synthase (papA), from Streptomyces
CC		venezuelae. papA participates in the biosynthesis pathway from chorismic
CC		acid into p-aminophenylpyruvic acid, and so the papA gene can be used to
CC		produce the transformant of the present invention. The transformant can
CC		be used to produce metabolites for application in pharmaceuticals,
CC		veterinary drugs and agrochemicals
XX		
SQ		Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other;
		Alignment Scores:
		Pred. No.: 4,65e-245 Length: 2061
		Score: 3573.00 Matches: 686
		Percent Similarity: 100.00% Conservativeness: 0
		Best Local Similarity: 100.00% Mismatches: 0
		Query Match: 100.00% Indels: 0
		Gaps: 0
		US-10-089-514-2 (1-686) X AAP86409 (1-2061)
OY		
Db	1	MetAgtHrLeuLeuIleAspSntlyrAspSerPheThHisAnLeuPheGlnTyrIle 20
Db	1	ATGCCGACCGCTTCGTGCACAACTCAGACTCCTTGACCACCAACCTGGTCCAGTCAATC 60
OY	21	GIYGIuaIaThGlVgInPrOpVaIVaIVaIProASnaSPMaIaSPTrPsArAgIeu 40
Db	61	GGCGAGGCCACCGGGGAACCCCCGTCGTGTGCCAACGACGCCCATGTGTGGCGCTG 120
OY	41	ProValGIuASPpheAspaIaIleValIaISerProGIyProGIySerProAsPARAgIU 60
Db	121	CCCGTCGAGGACTTCGACCGCATCGTCGTGTCCCGGGGCCCGGACGCCCGACCGGGAA 180
OY	61	ArgaspPheGIyIleSerarArgAlaIleThrAspSerCIyleuProValIeuGIyVal 80
Db	181	CGGGACTTCGGAAATCAGCGCGCGGGGATCACCGACAGCGGCGCTGCCTCCGCGCTC 240
OY	81	CysLeuGIyHiSGInGIyIleIaAGInIeuPheGIyGIyThTaIGIyleuIaIaProGU 100
Db	241	TGCCCTCGGCCACCGAGGCATCGCCAGCTTCCTCGGGGAAACCTCGGCGCTCGCCCGGAA 300
OY	101	PromethiEGlYArgValISergIuValArgHisETHGlYGIuSPaPyIaPheaGGIyleu 120
Db	301	CCCATTGACGCGCGGGGCTTCGAGGTGGCGACACCGGCGAGGACGTTCTTCGGGGGCTC 360
OY	121	ProSerProPheThraIaValaIArgTyrHISSerIeuIaIaIaThRaSPleuProAsPGIU 140
Db	361	CCCTCGCGGTTACCGCGCGGTGCTACACTCCTCGCGGCCACCGACCTCCCGACGAG 420
OY	141	LeuGIuProIeuIaITrpSerAspaPGIyValIaMecGIyleuArgHisArgGIuIyS 160
Db	421	CTCGAACCCCTCGCGCGGACGACCGGGGTGTGTATGGGCTTCGGCGACCGCGAGAG 480

QY	161	ProLeuTrpGlyValGlnPheHisProGluSerTLeGlySerAspPheGlyValGluIle	180
Db	481	CCGGCTGGGGCGCTCCAGTTCACCCCGAGTCCATCCGACGCGACTTCGCGCGGAGATC	540
QY	181	MetLAsnAspPheAAspAspLeuAlaLeuAlaHisHisAspArgLAspAspGHisGlyValAsp	200
Db	541	ATGGCAACTTCCCGGACCTCGCCTCGCCACCAACGAGGACCGCGCCGACGGGGCCGAC	600
QY	201	SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg	220
Db	601	TCCCGCATCGAACTCCACGTGCGCGCGGTGACGTGCTGCCGGACGCCGAAGAGTACGC	660
QY	221	ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerSerValLeuGlu	240
Db	661	CGCGAGCTGCCCGCCGCGAGGAGCACACGTTCTGGCTGGACACAGCACTCGTCTCGTAA	720
QY	241	GlyValSerArgPheSerPheLeuGlyValAspAspArgGlyProLeuAlaGluTyrLeuThr	260
Db	721	GAGCGCTCGCGCTTCCTCTCTCTCGGGAGACACCGCGAGCCGCGCTCGCGAGTACTCAC	780
QY	261	TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgThr	280
Db	781	TACCGCGTCCGCGACGCGCTGCTTCGCTCCGCGGCTCCAGCGGACACACGACCCGAGC	840
QY	281	ArgArgProPheAsnTyrLeuGluGluGlnLeuGluValArgArgValProValAla	300
Db	841	CGGCGCCCTCTTCAACTACTACGAGAGACGCTCGAACGCGACGCGGTCCCGTGC	900
QY	301	ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys	320
Db	901	CCCGAATCGCCCTTCGAGTTCACCTCGGCTACGTGCGTACTCTCGGCTACGACTTAAG	960
QY	321	AlaGluThrThrThrGlyAspProAlaHisAspSerProHisProAspAlaAlaPheLeuPhe	340
Db	961	GCGAGACACACCGCGGACCCCGGCCACCGGTCCCGGACCCCGACCGCGCTCTCTTC	1020
QY	341	AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeuAlaLeuAsp	360
Db	1021	GCCGACGCGCGCATCGCCCTCGACCAACGAGAAAGCTGCTACTGCTGCGCCCTCGAC	1080
QY	361	ArgArgGlyHisAspAspArgValAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr	380
Db	1081	CGCGGGGCGCACGACGAGCGCGCCGCGCTCGTGGGGAGACGGCTCGACACTTCAC	1140
QY	381	GlyLeuAlaValArgAlaProAlaGlyProThrProAlaMetValPheGlyIleProGlu	400
Db	1141	GCGCTGCGCGTCCGCGCCCGGCGGACCGACCCCGCATGCTCTTCGCGATCCCGAG	1200
QY	401	AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys	420
Db	1201	GCGGCGGCGGCTTCGCGCCCTCGCGCGCGCGCACGACGACGACGACGACTTACTCAAG	1260
QY	421	ArgTlleAspGlyCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThr	440
Db	1261	CGCATTCGACGAGTCTCTCAAGAGATCCGCGAACCGCGAGTCTCGAGTCTGCTCAC	1320
QY	441	AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla	460
Db	1321	AACATGGTCACCGCGCACCGAGGCGACGCGCTCGCGCTTACTCCGCGCTCGCGCC	1380
QY	461	IleSerProValProTyrGlyValAlaLeuLeuGluPhePheProGluLeuSerValLeuSerAla	480
Db	1381	ATCAGCCCCCGTCCGATCGGCGCCCTGCTCGAGTTCCTCCGAACTGTGGTCTCAGCGCC	1440
QY	481	SerProGluArgPheLeuThrTlleGlyValaAspGlyGlyValGluSerLysProIleLys	500
Db	1441	TGCGCCGAGCGGTTCCTCACGATCGGCGCGACGCGCGCGCTCGAGTCCAAAGCCATCAAG	1500
QY	501	GlyThrArgProArgGlyGlyThrAlaGluGluAspGluValArgLeuArgAlaAspLeuAla	520
Db	1501	GAGACCCGCGCCCGGCGGCGACCGCGGAGAGGACGAGCGGCTCCGCGCGACTTGACC	1560


```
QY 521 G1YArgGluYAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
DB 1561 GGGCGGAGAGAGACCGGGCGAGAACTGATGCTCACTGGTCCGCAACGACTTC 1620
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
DB 1621 AACAGCGTTCGGCGATCGGCTCCGTCACGTGCGCCGCTTTGAGGTGAGACCTAC 1680
QY 561 AlaProValHisGluLeuValSerThrIleArgGlyValArgLeuArgProGlyThrSerThr 580
DB 1681 GCGCCGCTGCACCACTGCTGTGACCATCCGGGAGAGCTGGCGCCGACACGACACC 1740
QY 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPolysArg 600
DB 1741 GCGCGCTGCTACGCGCGCTTCCCGGCGCTCCATGACCGGCGGCCCAAGAGCGC 1800
QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
DB 1801 ACCATGGAGATCATCGACCGCTGAGAGAGCGCCCGGGCGCTACTCCGGGGCGCTC 1860
QY 621 GlyTyrPheAlaLeuSerGlyValAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
DB 1861 GGATGGTTCGCGCTTCAGCGCGCGCGCGGACCTGACATCGTCATCCGACCATCGTCTG 1920
QY 641 AlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGlu 660
DB 1921 GCGGACGGCGGAGCGGAGTTCCGCGCTCGCGGGCGCATCTGTCCCTTCGACCGAGAG 1980
QY 661 GluGluPheThrGluThrValIleValAlaArgAlaMetValThrAlaLeuAspGlySer 680
DB 1981 GAGGAGTTGACCGAGACCGGTGTAAGGCCGCGGCGCATGTGTACCGCCCTCGACGGGAGC 2040
QY 681 AlaValAlaGlyAlaArg 686
DB 2041 GCGGTGGCGGCGCGCCGA 2058

RESULT 2
AAL50180
ID AAL50180 standard; DNA; 2061 BP.
XX
AC AAL50180;
XX
DT 28-JAN-2003 (first entry)
XX
DE S venezuelae PF1022 substance gene #1.
XX
KW Substance PF1022; biosynthesis; transgenic; choriismic acid; anthelmintic;
KM phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
NN veterinary drug; gene; ds.
XX
OS Streptomyces venezuelae.
XX
FH Key
FH CDS 1. 2061
FT /*tag= a
FT /product= "PF1022 substance"
XX
XX WO20027244-A1.
XX
XX 03-OCT-2002:
XX
XX PD
XX PF 22-MAR-2002; 2002WO-JP002782.
XX
XX PR 22-MAR-2001; 2001JP-00082227.
XX
XX PA (MEIJ ) MEIJI SEIKA KAISHA LTD.
XX
XX Yanaí K, Sumida N, Watanabe M, Moriya T, Murakami T;
XX WPI; 2003-018934/01.
XX DR P-PSDB; AAO19563.
XX
XX Novel biosynthesis gene-transferred transformants for producing PF1022
```

```
PT substance derivatives by fermentation, as pharmaceuticals or veterinary
PT drugs with anthelmintic activity.
XX
XX Claim 13; Page 74-79; 116pp; Japanese.
XX
XX The present invention relates to transformants capable of producing
CC PF1022 substance derivatives. These were obtained by transferring a gene
CC participating in the biosynthesis pathway from choriismic acid to p-
CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-
CC requiring host derived from an organism producing the PF1022 substance.
CC The transformants are producing PF1022 substance derivatives by
CC fermentation, for use as pharmaceuticals or veterinary drugs. The present
CC sequence is a substance PF1022 coding sequence from Streptomyces
CC venezuelae
XX
SQ Sequence 2061 BP; 297 A; 842 C; 654 G; 268 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 4,65e-245 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 8 Gaps: 0

US-10-089-514-2 (1-686) x AAL50180 (1-2061)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
DB 1 ATGCGCAGCCTTCTGATCGACAACATCGACTCGTTCACCCACACCTGTTCCAGTACATC 60
QY 21 GlyGluAlaThrGlyGlnProProValValValProAsnAspAlaAspTTPSerArgLeu 40
DB 61 GGGAGAGCCACCGGGGCAACCCCGCTGTCGTCGCCCAAGAGCGGACATGTCGCGCTG 120
QY 41 ProValGluAspPheAspAlaIleValIleSerProGlyProGlySerProAspArgGlu 60
DB 121 CCGGTGAGAGACTTCGACCGCATCGTGTGTCGCCGCGGCGGACGCCCGGAGCGGAA 180
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleGlyVal 80
DB 181 CGGGACTTCGGAATGACCGCGCGGGGATCACCGACAGCGGCGCTGCTCCGCGCTC 240
QY 81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 241 TGCCTCGGCACACAGGCGATCGCCAGCTCTCGGGGAAACGTCGCGCTCCCGCGGAA 300
QY 101 ProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
DB 301 CCCATGCAAGCGCGGGTCTCCAGAGTGCAGCAACCGGCGAGAGCTTTCGCGGCGCTC 360
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspGlu 140
DB 361 CCTCGCGGTTTACCGCGCGTGGCGTACCATCCTCGGCGGCAACGACTCCCGGACGAG 420
QY 141 LeuGluProLeuAlaTTPSerAspAspGlyValIleMetGlyLeuArgHisArgGlyLeu 160
DB 421 CTCGAACCCCTCGCGTGCAGCAAGCGGGGTCTCATGAGGCGCTGCGGCGGCGAGAG 480
QY 161 ProLeuTTPGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIle 180
DB 481 CCGCTGGGGGCGTCCAGTTCCACCGGAGTCCATCGGAGGAGCTTCGCGCGGAGATC 540
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArgHisGlyAlaAsp 200
DB 541 ATGGCCAACTTCGCGGACCTCGCGCTCGGCGCAACCGGCGACGGGCGACGGGCGGAGC 600
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
DB 601 TCCCGGTAGAACTCCACCTGCGCGCGCTGCGACGTGCTCCGAGAGCGCCAGAGATACCC 660
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTTPLeuAspSerSerValIleGlu 240
```

Db 661 CGCGCTGCTGCCCGGCGAGGCGACCACTTGTGCTGACAGACTTCCTCTCGAA 720
 QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
 Db 721 GGGGCTCGGGCTTCTCTCTCTCTCTCGGGAGAGACCGGGCCGCTCGCGAGTACTCTCACC 780
 QY 261 TyrArgValAlaAspGlyValValaSerValArgGlySerAspGlyThrThrThrArgThr 280
 Db 781 TACCGGTGCGCGAGCGCGTCTCTCGCTCGGGCTCGGACCGGACACAGACCCGGAGG 840
 QY 281 ArgArgProPhePheSerTyrLeuGluGluGluLeuGluArgArgValProValAla 300
 Db 841 CGGCGCCCTCTTCTTCAACTACCTGAGAGAGACAGCTGAAAGCGGAGCGGCTCCCGTCGCG 900
 QY 301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320
 Db 901 CCGGAAGTGGCTTTCAGATTCAACTCGGCTACGCTAGGCTACCTCGGCTTACGAGCTGAG 960
 QY 321 AlaGluThrThrGlyAspProAlaAlaIleArgSerProHisProAspAlaAlaPheLeuPhe 340
 Db 961 GCGGAGACCAACCGGCAACCCGCGCACCGGATCCCGCACCCGACCGCGGCTTCTCTTC 1020
 QY 341 AlaAspArgAlaAlaAlaLeuAspHisGluGluGlyCysGlyTyrLeuLeuAlaLeuAsp 360
 Db 1021 GCGGACCGGCGCATCGCTCTCGACACAGAGAGGCTGCTGCTTACCTGCTGCGCTCTGAC 1080
 QY 361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrLeuArgGlyThrAlaGluThrLeuThr 380
 Db 1081 CGCGGGGGCCACGACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1140
 QY 381 GlyLeuAlaValaArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
 Db 1141 GGGCTGCGCGTCCGCGCCCGCGCGAGCGACCCCGGCATGCTTCCGGAGTCCCGGAG 1200
 QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLys 420
 Db 1201 GCGCGCGCGCTTCCGCGCCCGCTGCGCGCGCGCACGACGACGCTTACCTCTCAAG 1260
 QY 421 ArgIleAspGluCysLeuLysGluIleArgAsnGlyGlySerTyrGluIleCysLeuThr 440
 Db 1261 CGCATCGACGAGTCCCTCAAGAGATCCGCAAGCGGAGTCTGATCCGATCTGCTGAC 1320
 QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
 Db 1321 AACATGTCACCGCGCGACGAGCGAGCGGCTCGCTACTCCGGCTGCGCGCC 1380
 QY 461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
 Db 1381 ATCAGCCCGTCCGTAACGCGCGCTGCTGAGTTCCTCCGAACTGTGCTGACCGCC 1440
 QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLys 500
 Db 1441 TCGCCGAGCGGTCTCTCAAGATCGGCGCGAGCGGCGGCTGAGTCCAAAGCCCATCAAG 1500
 QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
 Db 1501 GGGACCCGCGCCCGGCGGCGACCGCGAGAGAGAGAGGCGCTCGCGCGGACCTGGCC 1560
 QY 521 GlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAlaPhe 540
 Db 1561 GGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1620
 QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluTyrTyr 560
 Db 1621 AACAGGCTCGCGCGATCGGCTCCGTCACAGTGCCTCCGCTTCCAGGTGAGAGACTTAC 1680
 QY 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
 Db 1681 GGGCGCGTGCACAGCTGTGTGTGACCATCCGGGAGAGCGGTGGCGCGGACACAGACAC 1740
 QY 581 AlaAlaCysValaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600
 Db 1741 GCGGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1800

QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeu 620
 Db 1801 ACCATGAGATCATCGACCGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1860
 QY 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
 Db 1861 GGATGGTTCGCCCTCAGCGCGCGCCGCGACCTCAGCATGTCATCGCACCATCGTGTG 1920
 QY 641 AlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGluGlu 660
 Db 1921 GCGGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1980
 QY 661 GluGluPheThrGlyThrValValLysAlaArgAlaMetValThrAlaLeuAspGlySer 680
 Db 1981 GAGAGTTCACGAGACCGTGTAAAGCCCGCGCATGTATCCGCTTCAACGCGACG 2040
 QY 681 AlaValAlaGlyAlaArg 686
 Db 2041 GCGGTGCGGCGCGCGA 2058
 RESULT 3
 ABZ69799 standard; DNA; 3305 BP.
 XX ABZ69799;
 AC ABZ69799;
 XX
 DT 08-APR-2003 (first entry)
 XX
 DE plasmid papABC.
 XX
 KW DHFR; translation; orthogonal tRNA: O-tRNA; O-RS; CAT; TYRS;
 KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
 KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
 KW biosynthesis; p-aminophenylalanine; paf; chorismate.
 OS Synthetic.
 XX
 PN WO200285923-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002WO-US012465.
 XX
 PR 19-APR-2001; 2001US-0285030P.
 PR 06-FEB-2002; 2002US-0355514P.
 XX
 PA (SCRI) SCRIP9 RES INST.
 XX
 PI Schultz P, Wang L, Anderson JC, Chin JMK, Liu DR, Magliery TJ;
 PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
 DR WPI, 2003-120430/11.
 XX
 XX Composition useful for producing protein comprising unnatural amino acid,
 PT has translation system comprising orthogonal tRNA and orthogonal
 PT aminoacyl tRNA synthetase.
 XX
 PS Example 4; Page 127-128; 188pp; English.
 XX
 CC The invention relates to a novel composition comprising a translation
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
 CC -tRNA with at least one unnatural amino acid in the translation system
 CC and the O-tRNA recognises at least one selector codon. A composition of
 CC the invention is useful for producing at least one protein comprising at
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
 CC dihydrofolate reductase comprising CHOCH His6tag. The unnatural amino
 CC acid is provided exogenously. The translation system is a cell and the
 CC unnatural amino acid is biosynthesised by the cell. The present sequence
 CC represents a plasmid containing the individual genes papABC that encode

CC the enzymes used to carry out the conversion of chorismate to the
XX unnatural amino acid p-aminophenylalanine (pAF)
SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.38e-213	Length:	3305
Score:	3127.50	Matches:	618
Percent Similarity:	91.40%	Conservative:	9
Best Local Similarity:	90.09%	Mismatches:	42
Query Match:	87.53%	Indels:	17
DB:	8	Gaps:	3

US-10-089-514-2 (1-686) x ABZ69799 (1-3305)

```
QY 1 MetArgThrIleuLeuIleAspAsnTYrAspSerPheThrHisAsnLeuPheGlnTYrIle 20
DB 1 ATGGGCAAGCTTGTGATCGAACATACGACTCGTTCCACGAACTGTTCCAGTATCATC 60
QY 21 G1YGLAAlaThrG1YglnProValIValIProAsnAspAlaAspTpsSerArgLeu 40
DB 61 GGGAGAGCCACCGGGACGCCCC--GTCTGCCCCAACGACGCCGACTGTGCGGCTG 117
QY 41 ProValGluAspPheAspAlaIleValIValIserProG1YProG1YserProAspArgGlu 60
DB 118 CCCCTCGAGAGACTTCGACCGCGATCGTGTCTCCCGGGCCCCGGACGCCGACCGGGAA 177
QY 61 ArgAspPheG1YIleSerArgArgAlaIleThrAspSerG1YLeuProValIleuG1YVal 80
DB 178 CGGAGACTTCGGAGTACAGCCCGCGGCGATCACCGACAGCGGCTCGCTCCGCGGCTC 237
QY 81 CyeLeuG1YH1sgInG1YIleAlaGlnLeuPheG1Yg1YThrValIglYLeuAlaProGlu 100
DB 238 TGGCTGGGCCACAGGGGATGCCCCAG-----CTCTGGCGGAA 276
QY 101 ProMetHisG1YArgValIserGluValIArgHisThrG1YgluAspValPheArgG1YLeu 120
DB 277 CCCATGACGGCGGGGTCTCCGAGGTGCGGACACCGGCGAGGACGTCTTCGCGGGCTC 336
QY 121 ProSerProPheThrAlaValIArgTYrHisSerLeuAlaIleThrAspLeuProAspGlu 140
DB 337 CCTCCCGCTTCCACCGCGCTGCGCTACCTCCCTGCGCCGACCGACCTCCCGACGAG 396
QY 141 LeuGluProLeuAlaIleTpsSerAspAspGluValIValIleMetG1YLeuArgHisArgGlu 160
DB 397 CTGGAACCTCTCCCTGAGAGCAGAGCGGTGCTCATGGGCTTGCGGACCCCGAGAG 456
QY 161 ProLeuTpsG1YValGlnPheHisProGluSerIleG1YSerAspPheG1YArgGluIle 180
DB 457 CGGCTATGGGCGTCCAGTCCACCGGAGTCCATCGGACGCACTTCGCGCGGAGATC 516
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArgHisG1YAlaAsp 200
DB 517 ATGGCAACATTCGCGCACTCGCTCGCCACACCGGAGCATGTGCGGACGCGGCGAG 576
QY 201 SerProTYrGluLeuHisValIArgArgValIAspValIleuProAspAlaGluGluValIArg 220
DB 577 TGGGCTACCAACTCCACGTCGCGCGGTGACGTGTGCGGACCGCGGAAGGATACGC 636
QY 221 ArgG1YCysLeuProG1YglnG1YThrThrPheTpsAspSerSerSerValIleuGlu 240
DB 637 CGGCGTGGCGCCCGGCGGACGCGCAGCTTGTGGTGAACAGACGCTCCGCTCGAA 696
QY 241 G1YAlaSerArgPheSerPheLeuG1YAspAspArgG1YProLeuAlaGluTYrLeuThr 260
DB 697 GGGGCTCGCTTCCTTCCTTCGCGGAGAGACCGGCGCGCTCGCGGATCACTCAC 756
QY 261 TYrArgValAlaAspG1YValIleSerValIArgG1YSerAspG1YThrThrArgGln 280
DB 757 TACGGGTGCGCCAGCGCGTGTCTCGGTCGCGGCTCCGACGCGACAGACCGCGGAG 816
QY 281 ArgArgProPhePheAsnTYrLeuGluGluGluGluArgArgValIProValAla 300
DB 817 GCGGCAACCTCTTTCAGTCACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 876
QY 301 ProGluLeuProPheGluPheAsnLeuG1YTYrValG1YTYrLeuG1YTYrGluLeuIys 320
DB 877 CCCGACCTGCTTTCAGTTCACCTCGGTACGTGGCTTACCTGAGCTACGAGTGAAG 936
QY 321 AlaGluThrThrG1YAspProAlaHisIAspSerProHisProAspAlaAlaPheLeuPhe 340
DB 937 GCGGAGACACACCGGCGACCCCGGAGTACCGGCCCGACCCGACCGCGGCTTCCTTC 996
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnG1YCysGTYrTYrLeuLeuAlaLeuAsp 360
DB 997 GCGGACCGGCGCATCCGCTCGACACAGAGAGGCTGCTTACCTGTGCGGCTTCGAC 1056
QY 361 ArgArgG1YHisAspAspG1YAlaArgAlaIleTpsLeuArgGluThrAlaGluThrLeuThr 380
DB 1057 CGCGGAGCGCACACAGCGCGCTGCTGCTGCGGAGAGCGGCGGAGACCTTCACC 1116
QY 381 G1YLeuAlaValIArgAlaProAlaGluProThrProAlaMetValPheG1YIleProGlu 400
DB 1117 GGCCTGGCGCTCGGCTCGGCGGAGCGGACCCCGCATGCTTTCGGGCTCCCGAG 1176
QY 401 AlaAlaAlaG1YpHeG1YProLeuAlaArgAlaArgHisAspIysAspAlaTYrLeuIys 420
DB 1177 GCGCGGCGGCTTCGCGCCCTGCTCGGAGCGGACGACAGAGAGAGCGC----- 1227
QY 421 ArgIleAspGluCysLeuIysGluIleArgHisGlnIleG1YserTYrGluIleCysLeuThr 440
DB 1228 -----TCGGGCTCGGACAGGAGAGTGTGACAGATCTGCTGAGCC 1269
QY 441 AsnMetValIleThrAlaProThrGluAlaThrAlaLeuProLeuTYrSerAlaLeuArgAla 460
DB 1270 AACATGATACCGCGCGGACCGAGGCGAGCGGCTCCGCTACTCTCGGCTGCGCCGC 1329
QY 461 ILeSerProValIProTYrG1YAlaLeuLeuGluPheProGluLeuSerValIleuSerAla 480
DB 1330 ATCAGCCCGTCCGCTGCGGCTCGCTCGAGTTCGCGAGCTCGGCTCAGCGCC 1389
QY 481 SerProGluArgPheLeuThrIleG1YAlaAspG1YValIglYserIysProIleIys 500
DB 1390 TCGCCCGAGCGTTCCTCAAGATCGGCGGAGAGCGGCGGTGTCAGATCCAGCCATCAAG 1449
QY 501 G1YThrArgProArgG1Yg1YThrAlaGluGluIleuAspGluArgLeuArgAlaAspLeuAla 520
DB 1450 GGGACCGCTCCCGGGGCGCACCGGCGAGAGAGAGAGAGGCTCCGCGGACCTGAGCC 1509
QY 521 G1YArgGluIysAspArgAlaGluAsnLeuMetIleValIAspLeuValIArgAsnAspLeu 540
DB 1510 GGGCGGAGAGAGAGAGCGGCGGAGACCTGATGATGTGCACTTGTCGCGAACGACCTC 1569
QY 541 AsnSerValIcysAlaIleG1YserValIHisValIProArgLeuPheGluValIgluThrTYr 560
DB 1570 AACAGGCTGCGCGCATCGGCTCCGTCACAGTGCCTCCGCTTCGAGGTGGAGAGACTTC 1629
QY 561 AlaProValHisGlnLeuValIserThrIleArgG1YArgLeuAspProG1YThrSerThr 580
DB 1630 GGGCGCTGACCACTGATGTGTGACCATCGGAGAGCGCTGCGGCGCGGACACACACC 1689
QY 581 AlaAlaCysValIArgAlaAlaPheProG1Yg1YSerMetThrG1YAlaProIysIysArg 600
DB 1690 GCGCGCTGCTAGCGGCGGCTTCCCGGCGGCTCATACCGGCGGCGCGCAAGAGGCA 1749
QY 601 ThrMetGluIleIleAspArgLeuGluGluGluProArgG1YValTYrSerG1YAlaLeu 620
DB 1750 CCCATGAGATATCAACGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1809
QY 621 G1YTPPheAlaLeuSerG1YAlaIleAspLeuSerIleValIleArgThrIleValIleu 640
DB 1810 GGAATGTTGCTTCAGCGGCGCGCGGACCTTCAGATGTGATCCGACCAATCGGTG 1869
QY 641 AlaAspG1YAlaGluPheG1YValIglYAlaIleValSerLeuSerAspGlnGlu 660
DB 1870 GCGGACGCGCGGCGGAGTTCGGGCTCGGCGGCGGAGTGTCTCTCCGACAGAGAG 1929
```

Oy	661	GIUcLUphEmhrctLuhrVaiValLYsAlaAgaAlmeVatLmraLeaAsGlySer	680
Db	1930	GAGAGATTCAAGGAGACCGTGTGCAAGGCCCGCATGTCAACCGCTTCAGCGGAGC	1989
Oy	681	AlaValAlaGlyAlaAArg	686
Db	1990	GCACTGGCGGCGCCCGCA	2007
RESULT 4			
ID	ABZ69798	standard; DNA; 12391 BP.	
XX	ABZ69798	
AC	ABZ69798;		
XX	08-APR-2003	(first entry)	
DT			
XX		Plasmid plaec-papabc.	
DE			
XX		DHER; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TYRS;	
XX		orthogonal aminoacyl tRNA synthetase; unnatural amino acid;	
KW		chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;	
XX		biosynthesis; p-aminophenylalanine; pAF.	
XX		Synthetic.	
OS			
XX		WO200285923-A2.	
PN			
XX		31-OCT-2002.	
PD			
XX		19-APR-2002; 2002MO-US012465.	
PF			
XX		19-APR-2001; 2001US-0285030P.	
PR		06-FEB-2002; 2002US-0355514P.	
XX			
XX		(SCRI) SCRIPPS RES INST.	
PA			
XX		Schultz P, Wang U, Anderson JC, Chin JWK, Liu DR, Magliery TU;	
P1		Meggere EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;	
PI			
XX		WPI; 2003-120430/11.	
DR			
XX		Composition useful for producing protein comprising unnatural amino acid.	
PT		has translation system comprising orthogonal tRNA and orthogonal	
P7		aminoacyl tRNA synthetase.	
XX			
XX		Example 4; Page 124-127; 188bp; English.	
PS			
XX		The invention relates to a novel composition comprising a translation	
CC		system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl	
CC		tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O-	
CC		tRNA with at least one unnatural amino acid in the translation system	
CC		and the O-tRNA recognises at least one selector codon. A composition of	
CC		the invention is useful for producing at least one protein comprising at	
CC		least one unnatural amino acid. The protein is the Asp127AG mutant of	
CC		chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse	
CC		dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse	
CC		dihydrofolate reductase comprising CHOCH His6tag. The unnatural amino	
CC		acid is provided exogenously. The translation system is a cell and the	
CC		unnatural amino acid is biosynthesised by the cell. The present sequence	
CC		represents a plasmid for use in the biosynthesis of p-aminophenylalanine	
CC		(pAF) in vivo	
XX			
SO		Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;	
Alignment Scores:			
Pred. No.:	2.06e-212	Length:	12391
Score:	3127.50	Matches:	618
Percent Similarity:	91.40%	Conservative:	9
Best Local Similarity:	90.09%	Mismatches:	42
Query Match:	87.53%	Indels:	17
DB:	8	Gaps:	3

US-10-089-514-2 (1-686) x ABZ69798 (1-12391)

OY		1	MetArgThrLeuLeuLleAspAsnTyraSpSerPheThriSaSnLeuPheGlnTyrlle	20
Dd		24	ATGGCAACGCTTGTGATGCACAATTAGACTCGTTACCAGAACCTGTTCAGATCATC	83
OY		21	GlyGluAlaThrGlyGlnProProValValValProAsnAspAlaAspTrpSerArgLeu	40
Dd		84	GCGAGGCGCACCGGAGCCCCC--GTCGTGCCAAGCAGCCGATGTGTCCGGCTG	140
OY		41	ProValGluAspPheAspAlaIleValValSerProGlySerProAspArgGlu	60
Dd		141	CCCCCGAGGACTTCAGACGCATGTGTGTCCC CGGCCCCCGGACGCCCCGACCGGAA	200
OY		61	ArgAspPheGlyIleSerArgAlaIleThrAspSeriGlyLeuProValLeuGlyVal	80
Dd		201	CGGAGCTTCGGAGTACGCCCGCGGCGATCAACGACAGCGGCTGTCCGCTCCGGGCTC	260
OY		81	CysLeuGlyHisGlnGlyTlalaGlnLeuPheGlyThrValGlyLeuAlaProGlu	100
Dd		261	TGCCTCGGCGCACGAGGCATGCCAG-----CTCTCGCGCGAA	299
OY		101	ProMetHisGlyArgValSerGluValAlaArgHisThrGlyGluAspValPheArgGlyLeu	120
Dd		300	CCCATGACAGGCGGGGTCCGAAGTGCGGACACCGCGAGACAGCTTCCGGGGCTC	359
OY		121	ProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAspGlu	140
Dd		360	CCCTCGCGGTTCACCGCGCGGTGGCTACACCTCCGTGGCGCGCACGACTCCGACGAG	419
OY		141	LeuGluProLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgHisArgGlyLys	160
Dd		420	CTCGAACCCCTCGGCTGGAGGACGACGCGCTCGTCA TGGGCTGCGGACCGGAGAG	479
OY		161	ProLeuTrpGlyValGlnPheHisProGlnSerIleGlySerAspPheGlyArgGluIle	180
Dd		480	CCGCGTAGGGCGGTCCAGTCCACCGGAGTCCATCGCACGACTTCGGCGGGAGATC	539
OY		181	MetaAsnPeaArgAspLeuAlaLeuAlaHisHisArgAlaArgHisGlyValaAsp	200
Dd		540	ATTGGCAACTTCCGCGACTCGCCCTCGCCACACCGGCGACGTCCGACGCGGCGAC	599
OY		201	SerProTyrgluLeuHisValaArgArgValaAspValLeuProAspAlaGluValaArg	220
Dd		600	TGGGGCTACGAATCCACGTGCGCGCGGTCCAGCTGTCGCGGACGCGGAAGATGACG	659
OY		221	ArgGlyCysLeuProGlyGlnGlyThrThrPheTrpLeuAspSerSerValLeuGlu	240
Dd		660	GCGCGTGGCTGCGCGCGGAGGGGCGCACGTTCTGGCTGACACAGCTCGCTCTGAA	719
OY		241	GlyAlaSerArgPheSerPheLeuGlyValAspAspArgGlyProLeuAlaGluTyrlleThr	260
Dd		720	GGCGCTTGGCGTCTCTCTCTCTGGGACGACCGCGGCCGCTCGCGGACTACCC	779
OY		261	TyrrArgValaIaAspGlyValaValaSerValaArgGlySerAspGlyThrThrArgThr	280
Dd		780	TACCGCGTCCGCGACGCGGTGTCTCGTCCGCGGCTCCGACGCGACACAGACCCGGGAC	839
OY		281	ArgArgProPhePheAsnTyrlleGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	300
Dd		840	GCGGCGACCCCTTTCAGCTACTCGAGAGGACGACTCGAACCGCGCGGGTCCGCTGCC	899
OY		301	ProGluLeuProPheGluPheAsnLeuGlyTyrTralGlyTyrlleGlnGlyTyrgluLeuLys	320
Dd		900	CCCGACCTGCGCTTGAATTACCTCGGCTACCTCGGCTACCTCGGCTACGACTGAAG	959
OY		321	AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe	340
Dd		960	GCGAGACACACCGGCGACCCCGGAGTACCGGCCCCCGACCCCGACGCCGCTCTCTTC	1019
OY		341	AlaAspArgAlaIleAlaLeuAspHisGlnGlnGlyCysCysTyrlleLeuAlaLeuAsp	360
Dd		1020	GCGACCGCGCATCGCTTGCATCCACACGAGGAGGCTGTCTCTACCTGCTGGCTGGAC	1079

```

QY 361 ArgArgGlyHHisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr 380
DB 1080 CGCGGGGGCCACGACGAGGGGCCCGCTGGCTGGAGAGAGCGCCGAGACCTTACCC 1139
QY 381 GlyLeuAlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
DB 1140 GGGCTGGCCGCTCCGCGCCGAGCGCCGAGCCCGCCGACATGGCTTCCGGGTCCTCCGAG 1199
QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspIleAspAlaTrpLeuLeu 420
DB 1200 GCGCGGGCGGCTTCCGCGCCCGCTGGCTGGCGACGACGACGACGACGACGACGACGACG 1250
QY 421 ArgIleAspGluLeuLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 440
DB 1251 -----TCGGCGCTCCGACGCGCGAGTCGTCGAGACTTGCCTTAC 1292
QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTrpSerAlaLeuArgAla 460
DB 1293 AACATGGTCAACCGCGCGACGAGCGAGCGCGCTGGCTTACCTCCGCGCTGGCGCGC 1352
QY 461 IleSerProValProTrpGlyAlaLeuLeuGluPheProGluLeuSerValIleuSerAla 480
DB 1353 ATCAGCGCCCGCTCCGCTGGCGCGCTGCTGAGTTCGCCGAGCTGTGCTCAGCGCC 1412
QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerIleuProIleuLeu 500
DB 1413 TCGCCGAGCGGCTTCTTCCGATCGGCGCGAGCGCGCGCGCTCGAGTCCAGGCCATCAG 1472
QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
DB 1473 GGGACCGCGCCCGCGGGCGACCGCGCGAGAGAGAGAGCGCGCTCCGCGCGACCTGGCC 1532
QY 521 GlyArgGluLeuAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
DB 1533 GCGCGGAGAGAGAGCGCGCGCGAGAACCTGATGTCGTCGACCTGGTCCGACAGACCTC 1592
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValIleuThrTrp 560
DB 1593 AACAGGCTGCGCGCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCCGCTCC 1652
QY 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
DB 1653 GCGCGCGTGCACAGCTGCTGTCACATCCGGGAGCGCTGGCGCGCGCGCGCGCGCGCGCG 1712
QY 581 AlaAlaCysValAlaArgAlaAlaPheProGlyGlySerMetThrGlyAlaProGlyLeuArg 600
DB 1713 GCGCGCTGCGTACGCGCGCGCTTCCCGCGCGCTCCATGACGCGCGCGCGCGCGCGAG 1772
QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTrpSerGlyAlaLeu 620
DB 1773 CCCATGAGATCTCATCCGCTCGAGAGAGCGCGCGCGCGCTTACCCCGCGCGCTC 1832
QY 621 GlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
DB 1833 GATGTTGGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1892
QY 641 AlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGluGlu 660
DB 1893 GCGGACGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1952
QY 661 GluGluPheThrGluThrValValValValAlaArgAlaMetValThrAlaLeuAspGlySer 680
DB 1953 GAGGAGTTACGAGACCGGTGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2012
QY 681 AlaValAlaGlyValAlaArg 686
DB 2013 GCAGTGCGGCGCGACGA 2030

```

RESULT 5
 AAH66076
 ID
 AAH66076 standard; DNA; 1860 BP.

```

AC AAH66076;
XX
DT 26-SEP-2001 (first entry)
XX
DE C glutamicum coding sequence fragment SEQ ID NO: 1111.
XX
KW Corynebacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
XX
PN EP108790-A2.
XX
PD 20-JUN-2001.
XX
PF 18-DEC-2000; 2000EP-00127688.
XX
PR 16-DEC-1999; 99GP-00377484.
PR 07-APR-2000; 2000UP-00159162.
PR 03-AUG-2000; 2000UP-00280988.
XX
PA (KYOW ) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI, 2001-376931/40.
XX
DR P-PSDB; AAG90857.
XX
PT Novel polynucleotides derived from Corynebacterium bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analyzing
XX expression profile or pattern of a gene and identifying homologous gene.
XX
PS Claim 8; SEQ ID NO 1111; 246bp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Corynebacterium bacteria Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of Corynebacterium bacteria, measuring expression amount and analyzing
CC the expression profile or expression pattern of a gene derived from
CC Corynebacterium bacteria, and identifying a homologue of a gene derived from
CC Corynebacterium bacteria. Corynebacterium bacteria are useful for producing amino
CC acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
XX in electronic format directly from the European Patent Office
XX
SQ Sequence 1860 BP; 394 A; 478 C; 516 G; 472 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 1.61e-96 Length: 1860
Score: 1487.00 Matches: 328
Percent Similarity: 60.06% Conservative: 78
Best Local Similarity: 48.52% Mismatches: 204
Query Match: 41.62% Indels: 66
DB: Gaps: 13
XX
US-10-089-514-2 (1-686) x AAH66076 (1-1860)
QY 1 MetArgThrLeuLeuIleAspAsnTrpAspSerPheThrHisAsnLeuPheGluTrpIle 20
DB 1 ATGCGCGTTTATATATGATATATGATATCTTTCACGTTTAATCCGACCATATGG 60
QY 21 GlyGluAlaThrGlyGluProProValValValProAsnAspAlaAspTrpSerArgLeu 40
DB 61 GAAGAGTTTACGAGTACGACCTGTGGTGGCTTAATGATCAAGAATATGATGAGATG 120
QY 41 ProValGluAspPheAspAlaIleValAlaSerProGlyProGlySerProAspArgGlu 60
DB 121 CTT-----TTGACCGCGCTCATCTCTACCTGGCGCGCGCGCGCGCGCGCGCG 171
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGluVal 80

```

```

Db      172 GCTGATTTGGTATCTGTGCGAGGCGTCATTGACGCGTCCGATTTTGGGTGG 231
Qy      81 CysLeuGlyYHIsGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
Db      232 TGTTTAGCCACACAGGCGCATTTGGCTTGCGCTATGGCGGATGTTGATTTGGGCCACG 291
Qy      101 ProMetHisGlyYArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
Db      292 CCGGTCACCGGTGAGGTTTCGCGATCACCGCATGATGATTCAGGTTTATTTGACGACATC 351
Qy      121 ProSerProPheThrAlaValArgGlyYHISerLeuAlaIleThrAspLeuProAspGlu 140
Db      352 CTGAAACGTTTGAGCGCGGTGCTTATCTCGATGGCGCAACCCGCTTGGCGAGTGCA 411
Qy      141 LeuGluProLeuAlaIleTrpSerAspAspGlyValValMetGlyLeuArgHisArgGluLys 160
Db      412 TTGAAGCTACAGACTACACGCGATGATGTTGATATGATGCGATTCGACATGAAGTCTT 471
Qy      161 ProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180
Db      472 CCGCAGTGGGGTGTGCATTTTCATCCGGAATCTATTGGTGAACAATTCGGCCATCAGATC 531
Qy      181 MetAlaAsnPheArgSerLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyAlaAsp 200
Db      532 ATTAAAGACTCTCTTAAATTAGCGCGCACATATCGC----- 567
Qy      201 SerProTrpGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
Db      568 -----TGGCAATCTACAGAGAAACTATTCGCTCAGCGTTGATTCAGCAGCGGCTTTT 621
Qy      221 ArgGlyCysLeuProGluGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
Db      622 GAAACATTTCTTTCCTCCATTCCTCCCATGCTTTTGGCTGATGAT----- 666
Qy      241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTrpLeuThr 260
Db      667 ---GCCCAAGAACCGACTATCTTGGTGAATGCCAGCGGTCCTTCCACAGCACAAAACC 723
Qy      261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280
Db      724 CATATGTGCGCGAGGG----- 741
Qy      281 ArgArgProPhePheAsnTrpLeuGluGluGlnLeuGluArgArgValProValAla 300
Db      742 -----GATTTCTTCACTGCGCTAAAGAGGATCTGCGCGCAAC-----TCAGTTGGG 789
Qy      301 ProGluLeuProPheGluPheAsnLeuGlyTrpValGlyTrpLeuGlyTrpGluLeuLys 320
Db      790 CCGGTCGCA-----GGTTTCGTCCTTGGCTGGGTGTTACCTTGTTATGAGCTTAA 843
Qy      321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db      844 GCGGAAGCTGGCGCACGCGCTGCGCACACTTCGAGTCTTCGCGATCGCACCTCATTTT 903
Qy      341 AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTrpLeuLeuAlaLeuAsp 360
Db      904 GCGGATCGCGCCATCCAGTCGATGATCGATCAG-----CTTCGGTTGGCTGGGTTG--- 954
Qy      361 ArgArgGlyYHISAspAspGlyAlaArgAlaIleTrpLeuArgGluThrAlaGluThrLeuThr 380
Db      955 -----GGGGAGCAGAC-----GAGTGGTTTGAAGAAACCTCAAGAGACGTGCAT 999
Qy      381 GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
Db      1000 AATCTTGTCGCCCGCGGATACCTGCGTCC----- 1029
Qy      400 GluAlaAlaAlaGlyPheGlyProLeuAla---ArgAlaArgHisAspLysAspAlaTrp 418
Db      1030 -----GGACACCTCGCTTGGCAGTTTCGAGATTCCAAAGTGAAGTAT 1071
Qy      419 LeuLysArgTrpLeuAspGluCysLeuLysGluIleArgAsnGlyGluSerTrpGluIleCys 438

```

```

Db      1072 CTCGACAAATTCGACAGCCACGAGACTGATTACTCGCGCGGAATCGATGAATCTGC 1131
Qy      439 LeuThrAsnMetValThrAlaProThrGluAlaIleAlaLeuProLeuTrpSerAlaLeu 458
Db      1132 CTGACACAAACTTCAGGCGACACGTATGATGGCCCTCTGGCTGCTATCTAGACACTG 1191
Qy      459 ArgAlaIleSerProValProTrpGlyAlaLeuLeuGluPheProGluLeuSerValLeu 478
Db      1192 CGTGGGGCCCAATCCACCCGATATGATGCGTATCTTCAGCTGGGAGTACCTTATTTTG 1251
Qy      479 SerAlaSerProGluAspPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro 498
Db      1252 AGTTCCTCGCCGAGGGGTTTATCACCATTTGATTCGCGAGGGATGATGGAATCAAGCCC 1311
Qy      499 IleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluAspValArgAlaAsp 518
Db      1312 ATTAAAGGACACAGGCGCGGTGGCGAACAGCGCAAGAACCAAGAAATCATCTGCTGAG 1371
Qy      519 LeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
Db      1372 CTGCGCAGTAATCTTAAAGATGTCGACAAACTTGATGATGCTGGAATTTGGTCGCAAC 1431
Qy      539 AspLeuAsnSerValCysAlaAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
Db      1432 GACTTAGCCCGCGCGCTTGGCCACACAGTTAAACATCCAGCTTTTGCAGCTCGAA 1491
Qy      559 ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
Db      1492 ACTCAGCCACAGTCCACCACTTGTCAGCACCGTCTCGCAGATTTGGGGCCA---CGC 1548
Qy      579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Db      1549 AGTCGATTAAGTGGTGGCGCCAGCATTTCCCGGTGGTTTCATGATCGTGTGCCCAAG 1608
Qy      599 LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValLysSerGly 618
Db      1609 CTGCGCACCATGAGATCTCATATGATCGGAGGACGCTCTCGCGGATTTTACTCAGGT 1668
Qy      619 AlaLeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Db      1669 GCGTTGGGATATTTTTCCTCCACGCGCGAGTTGATCTCCATGTGTATCGAACTCTC 1728
Qy      639 ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAsp 658
Db      1729 GTCATCCAGAACATCAACGTGAGTACGAGAGTGGCGGTCATTCTTGCTGTCTGAT 1788
Qy      659 GlnGluGluGluPheThrGluThrValValLysAlaArgAlaMetVal 674
Db      1789 CCGAGGCTGAGTGGAGGAAATCCGCTTAAATCACGCGCTTGTCTG 1836

```

RESULT 6
 AAF72103
 ID AAF72103 standard; DNA; 1983 BP.
 XX
 AC AAF72103;
 XX
 DT 30-APR-2001 (first entry)
 XX
 DE Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:701.
 XX
 KW Corynebacterium glutamicum; metabolic pathway protein; MP protein;
 KW fine chemical production; microorganism; organic acid; nucleoside;
 KW nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;
 KW lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;
 KW carbohydrate; aromatic compound; cofactor; polyketide; enzyme; ds.
 XX
 OS Corynebacterium glutamicum.
 XX
 PN MO200100843-A2.
 XX
 PD 04-JAN-2001.
 XX
 PF 23-JUN-2000; 2000WO-1B000923.

XX	25-JUN-1999;	99US-0141031P.
PR	01-JUL-1999;	99DE-01030476.
PR	02-JUL-1999;	99US-0142101P.
PR	08-JUL-1999;	99DE-01031415.
PR	08-JUL-1999;	99DE-01031418.
PR	08-JUL-1999;	99DE-01031419.
PR	08-JUL-1999;	99DE-01031420.
PR	08-JUL-1999;	99DE-01031424.
PR	08-JUL-1999;	99DE-01031434.
PR	08-JUL-1999;	99DE-01031435.
PR	08-JUL-1999;	99DE-01031443.
PR	08-JUL-1999;	99DE-01031453.
PR	08-JUL-1999;	99DE-01031457.
PR	08-JUL-1999;	99DE-01031465.
PR	08-JUL-1999;	99DE-01031478.
PR	08-JUL-1999;	99DE-01031510.
PR	08-JUL-1999;	99DE-01031541.
PR	08-JUL-1999;	99DE-01031573.
PR	08-JUL-1999;	99DE-01031592.
PR	08-JUL-1999;	99DE-01031632.
PR	08-JUL-1999;	99DE-01031634.
PR	08-JUL-1999;	99DE-01031636.
PR	09-JUL-1999;	99DE-01032125.
PR	09-JUL-1999;	99DE-01032126.
PR	09-JUL-1999;	99DE-01032230.
PR	09-JUL-1999;	99DE-01032232.
PR	09-JUL-1999;	99DE-01032227.
PR	09-JUL-1999;	99DE-01032228.
PR	09-JUL-1999;	99DE-01032229.
PR	14-JUL-1999;	99DE-01032922.
PR	14-JUL-1999;	99DE-01032926.
PR	14-JUL-1999;	99DE-01032928.
PR	14-JUL-1999;	99DE-01033004.
PR	14-JUL-1999;	99DE-01033005.
PR	12-AUG-1999;	99US-0148613P.
PR	12-AUG-1999;	99DE-01040764.
PR	27-AUG-1999;	99DE-01040765.
PR	27-AUG-1999;	99DE-01040766.
PR	27-AUG-1999;	99DE-01040832.
PR	31-AUG-1999;	99DE-01041378.
PR	31-AUG-1999;	99DE-01041379.
PR	31-AUG-1999;	99DE-01041380.
PR	31-AUG-1999;	99DE-01041394.
PR	31-AUG-1999;	99DE-01041396.
PR	03-SEP-1999;	99DE-01042076.
PR	03-SEP-1999;	99DE-01042077.
PR	03-SEP-1999;	99DE-01042079.
PR	03-SEP-1999;	99DE-01042086.
PR	03-SEP-1999;	99DE-01042087.
PR	03-SEP-1999;	99DE-01042088.
PR	03-SEP-1999;	99DE-01042095.
PR	03-SEP-1999;	99DE-01042124.
PR	03-SEP-1999;	99DE-01042129.
PR	09-MAR-2000;	2000US-0187970P.
XX	(BADI) BASF AG.	
XX		
XX	Pompejus M, Kroegeer B, Schroeder H, Zeidler O, Haberhauer G;	
XX	WPI, 2001-137957/14.	
XX	P-PSDB; AAB79984.	
XX		
PT	Nucleic acids from Corynebacterium glutamicum encoding metabolic pathway	
PT	proteins, useful for producing fine chemicals in microorganisms,	
PT	including organic acids, nonproteinogenic amino acids, and purine and	
PT	pyrimidine bases.	
XX		
PS	Claim 3; Page 1130-1133; 1737P; English.	

[illegible]

OY		261	TyrArgValAlaIAspGlyValValSerValArgLysSerAspGlyThrThrArgThr	280
Db		824	CATAATTCGGCGAGGG-----	841
OY		281	ArgArgProPhePheAsnTyrlLeuGlulnleuGluAArgArgValProValAla	300
Db		842	-----GATTCTTCACTGGCTAAAGAGGATCTCGCCGCCAAC-----TCAGTTCCG	889
OY		301	ProGluLeuProPheGluPheAsnleuGlyTyrlValGlyTyrlLeuGlyTyrlnleuLys	320
Db		890	CCCCGTCAA-----GGTTTTCGTCTTGCTGGGTGGTGAACGTTTAGACTTAAGCTTAAA	943
OY		321	AlaGluThrThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe	340
Db		944	GCGGAAGCTGGCGCACCGGGCTGGCGCACCTTCGAGTCTTCGGAGTGGCGCACTCATTTTT	1003
OY		341	AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysGlyTyrlLeuValAlaLeuAsp	360
Db		1004	GCCGATCGGCATCGCATGCGATGGAGATCCGATCAG-----GTTCCGTTGCTGGCGTTG--	1054
OY		361	ArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr	380
Db		1055	-----GGGGAGCAGAGAC-----GAGTGGTTTGAAGAACAACATCAAAAGACTGCAT	1099
OY		381	GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro	399
Db		1100	AATCTTTCGCGCCCGCGGATACCTGCGTCC-----	1129
OY		400	GluAlaIalAlaGlyPheGlyProleuAla---ArgAlaArgHisAspLysAspAlaTyrl	418
Db		1130	-----GACACCTCGCTTTTGGCAGGTTCCAGATTCCAAGATGAGTAT	1171
OY		419	LeuLysArgGlaeAspGluCysleuLysGluIleArgAsnGlyLysSerTyrlGluIleCys	438
Db		1172	CTCGACAAATTCGAGAGACCAGAGCGGTATTTACTCGCGCGAGATGCTAAGAATCTGC	1231
OY		439	LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrlSerAlaLeu	458
Db		1232	CTGACCACAAAATTCCTCAGGGCACCCATGATGGCCCTTCGGTGCCTACTCTGACCTG	1291
OY		459	ArgAlaIleSerProValProTyrlGlyAlaLeuLeuGluPheProGluLeuSerValLeu	478
Db		1292	CGTGGGGCCAATCCCACCGCATATGTGTGCTGATCTTCAGCTGGGGGAAATCCTCAATTTTG	1351
OY		479	SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro	498
Db		1352	AGTTCCTCGCGGAGCGGTATCATCCATTGATTCCGCGAGGGTATGTGCATCAAGGCC	1411
OY		499	IleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAsp	518
Db		1412	ATTAAAGCACACGAGCCCGCTGGGGGGAACACGCGAAGAAAGACCAAAATCATTTGCTGAG	1471
OY		519	LeuAlaGlyArgGluLysAspArgAlaGluAsnMetIleValAspLeuValArgAsn	538
Db		1472	CTGCGCAGTAATCCTAAAGATGTGCAPAAAACTTAGATGATCGTGATTTGTCGCGAAC	1531
OY		539	AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu	558
Db		1532	GACTTAAGCCCGGGGCTTTGCCACCCACAGTTAAACAATCCMACCTTTTGACAGTGCMA	1591
OY		559	ThrTyrlAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyTyrl	578
Db		1592	ACCTTACGCCACAGTCCACCAACTTGTGCAGCACCGCTCTGCAGAGATTGGGGCCA---CGC	1648
OY		579	SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys	598
Db		1649	AGTCGATGTAGTGTGGTGGCGGACGATTTCCCGGTGGTTGATGACGTGGCCCAAG	1708
OY		599	LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrlSerGly	618
Db		1709	CTGGCACAATGAGATCATCATGATAGCTGGAGGACGCTCTCCGCGGATTTACTCAGGT	1768
OY		619	AlaLeuGlyTyrlPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle	638

Db	1769	GGCTTGGATATTTTCCCTCGACGGCGACAGTTGATCTTCACATGATCAGAACTTCTC	1828
Qy	639	ValLeuAlaEPGIgYlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAsp	658
Db	1829	GTCAATCCAGAACATCAACGTGAGTACGAGATGGCGGTGACATCTTGCTGTGTGAT	1888
Qy	659	GlnGluGluGluPheThrGluThrValValIleAlaArgalaMetVal	674
Db	1889	CCGAGGCGTCACTGAGGAGAAATCCGCGTTAAATCAGCGGCTTCGCTG	1936
RESULT 7			
ID	AAF71982	standard; DNA; 1983 BP.	
XX	AAF71982;		
XX	30-APR-2001	(first entry)	
DE	Corynebacterium glutamicum MP protein nucleotide sequence SEQ ID NO:459.		
XX			
KW	Corynebacterium glutamicum; metabolic pathway protein; MP protein;		
KW	fine chemical production; microorganism; organic acid; nucleoside;		
KW	nonproteinogenic amino acid; purine base; pyrimidine base; nucleotide;		
KM	lipid; saturated fatty acid; unsaturated fatty acid; diol; vitamin;		
KM	carbohydrate; aromatic compound; cofactor; polypeptide; enzyme; ds.		
XX			
OS	Corynebacterium glutamicum.		
XX			
PN	WO200100843-A2.		
XX			
PD	04-JAN-2001.		
XX			
PF	23-JUN-2000; 2000WO-IB000923.		
XX			
PR	25-JUN-1999;	99US-0141031P.	
PR	01-JUL-1999;	99DE-01030476.	
PR	02-JUL-1999;	99US-0142101P.	
PR	08-JUL-1999;	99DE-01031415.	
PR	08-JUL-1999;	99DE-01031418.	
PR	08-JUL-1999;	99DE-01031419.	
PR	08-JUL-1999;	99DE-01031420.	
PR	08-JUL-1999;	99DE-01031424.	
PR	08-JUL-1999;	99DE-01031428.	
PR	08-JUL-1999;	99DE-01031434.	
PR	08-JUL-1999;	99DE-01031435.	
PR	08-JUL-1999;	99DE-01031443.	
PR	08-JUL-1999;	99DE-01031453.	
PR	08-JUL-1999;	99DE-01031457.	
PR	08-JUL-1999;	99DE-01031465.	
PR	08-JUL-1999;	99DE-01031478.	
PR	08-JUL-1999;	99DE-01031510.	
PR	08-JUL-1999;	99DE-01031541.	
PR	08-JUL-1999;	99DE-01031573.	
PR	08-JUL-1999;	99DE-01031592.	
PR	08-JUL-1999;	99DE-01031632.	
PR	08-JUL-1999;	99DE-01031634.	
PR	08-JUL-1999;	99DE-01031636.	
PR	09-JUL-1999;	99DE-01032125.	
PR	09-JUL-1999;	99DE-01032126.	
PR	09-JUL-1999;	99DE-01032130.	
PR	09-JUL-1999;	99DE-01032186.	
PR	09-JUL-1999;	99DE-01032206.	
PR	09-JUL-1999;	99DE-01032227.	
PR	09-JUL-1999;	99DE-01032228.	
PR	09-JUL-1999;	99DE-01032229.	
PR	09-JUL-1999;	99DE-01032230.	
PR	14-JUL-1999;	99DE-01032922.	
PR	14-JUL-1999;	99DE-01032926.	
PR	14-JUL-1999;	99DE-01032928.	
PR	14-JUL-1999;	99DE-01033004.	
PR	14-JUL-1999;	99DE-01033005.	
PR	14-JUL-1999;	99DE-01033006.	


```

QY 161 ProLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180
DB 587 CCGGAGGAGGGGTGTCATTTTCATCCGGAATCTATGGTGGACAAATTCGGCAATCAGATC 646
QY 181 MetLalaSerPheAspLeuAlaLeuAlaHisIleSarGlaAArgAlaGHisGlyAlaAsp 200
DB 647 ATTAAGAACTTCTTAATTAGCCGCGCAATATCGC----- 682
QY 201 SerProTyrGluLeuHisValArgArgValAlaLeuProAspAlaGluGluValArg 220
DB 683 -----TGGCACTTCAGAGAAAATTAATTCCTGCTCAGCTGTATTCAGACGGGTTTTT 736
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
DB 737 GAAACATTTCTTGGCCATCTCTCCATGCTTTTGGCTCGATGAT----- 781
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
DB 782 ---GCCCAAGGAAACAGCTATCTTGATGATGCCAGCGTCTCTCGACGCAAAAACC 838
QY 261 TyrArgValAlaAspGlyValAlaSerValArgIleSerAspGlyThrThrThrArgThr 280
DB 839 CATATATCTCGCGCGAGGG----- 856
QY 281 ArgArgProPhePheAsnTyrLeuGluGluGlnLeuGluArgArgValProValAla 300
DB 857 -----GATTTCTTCACTGCTAAAGAGATTCGCCGCCAAC-----TCAGTTGCG 904
QY 301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys 320
DB 905 CCCGGTCAA-----GGTTTTCGTTGGTGGGTGGTTAGCTTGAAGCTTAA 958
QY 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
DB 959 GCGGAACTGGCGGCGCGGCTGCGCACATTCGAGTTCGCGATGCGCACTCATTTT 1018
QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuAlaLeuAsp 360
DB 1019 GCCGATGCGCGCATCGAGTGGAAATCGGATCAG-----GTTCGTTGCTGGCGTTG-- 1069
QY 361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrLeuArgGluThrAlaGluThrLeuThr 380
DB 1070 -----GGGAGCAGGAC-----GAGTGGTTGAAGAAACCATCAAGAGTCAT 1114
QY 381 GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
DB 1115 AATCTTGTCCGCCCGCGATACCTGCTCC----- 1144
QY 400 GluAlaIleAlaGlyPheGlyProLeuAla---ArgAlaArgHisAspLysAspAlaTyr 418
DB 1145 -----GGACACCTGCTTGGAGTTGGAATTCGAAATTCGAAAGATGAGTAT 1186
QY 419 LeuLysArgIleAspGlyCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCys 438
DB 1187 CTGCACAAAATTTGGCAGAGCCGAGACTGATTACTCGCGCAATGTATGAAATTCG 1246
QY 439 LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu 458
DB 1247 CTACCCCAAAACTTCAGGCGCACCATGATGCGCCCTGCTGCTGCTATTCAGCACTG 1306
QY 459 ArgAlaIleSerProValProTyrGlyValaLeuLeuGluPheProGluLeuSerValLeu 478
DB 1307 CGTGGGCGCCATCCACCGCATATGATGCTGATTTAGCTGGGGGATACCTCTATTTTG 1366
QY 479 SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro 498
DB 1367 AGTTCTCTCGCGGAGGCTTATCATCATGATTCGCGCAGGATATGCGAATCAAAACC 1426
QY 499 IleLysGlyThrArgProArgGlyGlyTyrAlaGluGluAspGluArgLeuArgAlaAsp 518
DB 1427 ATTAAGGACACAGCGCGCTGGCGCAAGCCGCAAGAAACCAAGAAATCATTTGCTGAG 1486

```

```

QY 519 LeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
DB 1487 CTGGCAGGTAATCTTAAGATCTGCGAAGAAATCTTGATGATCTGGATTTGGTCCGCAAC 1546
QY 539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
DB 1547 GACTTACGCCCGCGCGCTTTCCACACAGCTTAAACATCCAGCTTTTCAGCTCGAA 1606
QY 559 ThrThrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
DB 1607 ACCTAGCCCACTTCACCACTTGTACGACCGCTCTCGACAGTTGGGCGCA---CGC 1663
QY 579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
DB 1664 AGTCCAGTTAGAGGCGTCCCGCGCAGCATTCGCCGTTGCTGATGATGCTGCCCAAG 1723
QY 599 LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGly 618
DB 1724 CTGCGCACCATGGAGATCATCATGAGCTGGAGCGAGCTCTCGCGTATTTACTCAGGT 1783
QY 619 AlaLeuGlyThrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
DB 1784 GGCTTGGATATTTTCCCTCGACGCGCAGCTGATCTCCATGGTATCAGAACTTC 1843
QY 639 ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAsp 658
DB 1844 GTATTCAGAAACAATACGTCGAGTACGAGTGGGCGTGGCACTTCTGCTCTGTGAT 1903
QY 659 GlnGluGluGluPheThrGluThrValValIleValAlaArgAlaMetVal 674
DB 1904 CCGGAGGCTGAGTGGAGGAAATCCGCTTAATCAACGCGCTGTGCTG 1951

```

RESULT 9

AAH68527 standard; DNA; 349980 BP.

AAH68527;

26-SEP-2001 (first entry)

C glutamicum coding sequence fragment SEQ ID NO: 7062.

Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

organic acid synthesis; ds.

Corynebacterium glutamicum.

EPI108790-A2.

20-JUN-2001.

18-DEC-2000; 2000BP-00127688.

16-DEC-1999; 99P-00377484.

07-APR-2000; 2000JP-00159162.

03-AUG-2000; 2000JP-00280988.

(KYOWA) KYOWA HAKKO KOGYO KK.

Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

Tateishi N, Senoh A, Ikeda M, Ozaki A;

WPI; 2001-376931/40.

Novel polynucleotides derived from Coryneform bacteria, for identifying

mutation point of a gene, measuring expression of a gene, analyzing

expression profile or pattern of a gene and identifying homologous gene.

Disclosure; SEQ ID NO 7062; 246pp + Sequence Listing; English.

The present invention provides a number of nucleotide and protein sequences from the Coryneform bacterium Corynebacterium glutamicum. These are useful for identifying the mutation point of a gene derived from a

QY 619 AAlaLeuGIYrPheAlaLeuSerGIYAlaIaLaapLeuSerIleValIleArgThrIle 638
 Db 153689 GCGTTCGGATATATTTTCCCTCAGCGCGCAGCTGATCTCTCCATGGTATCAGAACTCTC 153748
 QY 639 ValIleuAlaAspGIYAlaIaLaGIuPheGIYValGIYGIYAlaIleValSerLeuSerAsp 658
 Db 153749 GTGATCCAGAACATCATCGTGGATACGAGATGGCGCGGCGCATCTTCTGCTGTGAT 153808
 QY 659 GIuGIuGIuGIuPheThrGIuThrValIleValIysAlaArgAlaMetVal 674
 Db 153809 CCGAGAGCTGATGGTGAGAGAAATCCGCGTTAATCATCAGCGCTCTGCTG 153856
 RESULT 10
 ACF67367_10
 Continuation (50 of 57) of ACF67367 from base 4900001 (Photorhabdus luminescens nucleoti
 WP Sequence split into 57 fragments LOCUS ACF67367 Accession ACF67367
 WP Fragment Name Begin End
 WP ACF67367_00 1 110000
 WP ACF67367_01 100001 210000
 WP ACF67367_02 200001 310000
 WP ACF67367_03 300001 410000
 WP ACF67367_04 400001 510000
 WP ACF67367_05 500001 610000
 WP ACF67367_06 600001 710000
 WP ACF67367_07 700001 810000
 WP ACF67367_08 800001 910000
 WP ACF67367_09 900001 1010000
 WP ACF67367_10 1000001 1110000
 WP ACF67367_11 1100001 1210000
 WP ACF67367_12 1200001 1310000
 WP ACF67367_13 1300001 1410000
 WP ACF67367_14 1400001 1510000
 WP ACF67367_15 1500001 1610000
 WP ACF67367_16 1600001 1710000
 WP ACF67367_17 1700001 1810000
 WP ACF67367_18 1800001 1910000
 WP ACF67367_19 1900001 2010000
 WP ACF67367_20 2000001 2110000
 WP ACF67367_21 2100001 2210000
 WP ACF67367_22 2200001 2310000
 WP ACF67367_23 2300001 2410000
 WP ACF67367_24 2400001 2510000
 WP ACF67367_25 2500001 2610000
 WP ACF67367_26 2600001 2710000
 WP ACF67367_27 2700001 2810000
 WP ACF67367_28 2800001 2910000
 WP ACF67367_29 2900001 3010000
 WP ACF67367_30 3000001 3110000
 WP ACF67367_31 3100001 3210000
 WP ACF67367_32 3200001 3310000
 WP ACF67367_33 3300001 3410000
 WP ACF67367_34 3400001 3510000
 WP ACF67367_35 3500001 3610000
 WP ACF67367_36 3600001 3710000
 WP ACF67367_37 3700001 3810000
 WP ACF67367_38 3800001 3910000
 WP ACF67367_39 3900001 4010000
 WP ACF67367_40 4000001 4110000
 WP ACF67367_41 4100001 4210000
 WP ACF67367_42 4200001 4310000
 WP ACF67367_43 4300001 4410000
 WP ACF67367_44 4400001 4510000
 WP ACF67367_45 4500001 4610000
 WP ACF67367_46 4600001 4710000
 WP ACF67367_47 4700001 4810000
 WP ACF67367_48 4800001 4910000
 WP ACF67367_49 4900001 5010000
 WP ACF67367_50 5000001 5110000
 WP ACF67367_51 5100001 5210000
 WP ACF67367_52 5200001 5310000
 WP ACF67367_53 5300001 5410000
 WP ACF67367_54 5400001 5510000
 WP ACF67367_55 5500001 5610000

[illegible]

QY 203 TyrGluLeuHISValArgArgValAspValLeuProAspAlaGluGluValArgArgGly 222
 Db 14551 CTCACCTCGAATTCGTGAATGTAAGAGATAGTCGATCCTGATCTTGTATTATCA 14610
 QY 223 CysLeuProGlyLysGluGlyThrThrPheTrpLeuAspSerSerValLeuGluGlyAla 242
 Db 14611 CGTTATGGCAATATACGATGCTTCTGCTGATAGTAAATCTGTATAGACCCAT 14670
 QY 243 SerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTrpLeuThrTrpArg 262
 Db 14671 GCCCGCTTCTCAATGATGGCGACCGGTAATGAACAGCTGCTGCAGACTTGAAATACGT 14730
 QY 263 ValAlaAspGlyValValSerValArgGlySerAspGlyThrThrTrpArgTrpArg 282
 Db 14731 GTACAGACTGAATTTGCGTTGATCGCTCGATGGCAGAGAGANT---GTTCAAGGA 14787
 QY 283 ProPhePheAsnTrpLeuGluGluGluLeuGluArgArgValProValAlaProGlu 302
 Db 14788 GATTTTTCCTCGTGATGTCTCAGTCTGATGTGGCAATTTCGACACCAAAATCG 14847
 QY 303 LeuProPheGluPheAsnLeuGlyTrpValGlyTrpLeuGlyTrpGluLeuValAlaGlu 322
 Db 14848 ATGCCATTGGATTATTAAGCGGGTTGTGTATTAGATATGATTAATTAAGCGCGCTG 14907
 QY 323 ThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAsp 342
 Db 14908 ACAATTGGCGACCAAAATATCTCTTCAAGAACATCTGATCCAGCTTAATATTACTCCG 14967
 QY 343 ArgAlaIleAlaLeuAspHisGluGluGlyCys-----CysTrpLeuAlaLeu 359
 Db 14968 CATTTTTTCCTCTTCATCATCATCAGCAGAGTAACTTATGAATGTTTATCACACCAAT 15027
 QY 360 AspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgLysThrAlaGluThrLeu 379
 Db 15028 GGGCGACCATCATAT-----TGGTCACCGGAACCTTCTGTTACCAAT 15069
 QY 380 ThrGlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
 Db 15070 ACA-----AAAAATACCGGTGAAGTATACCTGATCTTATTCCTGCGCGCTGTT 15117
 QY 400 GluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaAspHisAspLysAspAlaTrpLeu 419
 Db 15118 GATCAAAATAG-----ATTGTCTGGAATATGCGCGCTGAGAAATATATG 15162
 QY 420 LysArgIleAspGlyCysLeuLysGluLysArgAlaGluSerTrpGluIleCysLeu 439
 Db 15163 GATGGAATCCATTAATCATCCATCATCACAGATGAGTCCATGAAATCTGTCTG 15222
 QY 440 ThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTrpSerAlaLeuArg 459
 Db 15223 AGCAATAGGGCCCAAGATGACTTATTCAGACATCCATTCGACTTATCGCGCTATGCGT 15282
 QY 460 AlaIleSerProValProTrpGlyAlaLeuLeuGluPheProGluLeuSerValLeuSer 479
 Db 15283 CATGCCAGTCTCTACTTATGCGCTTATCTGCGCTGCGGATTTTCTGACTCACT 15342
 QY 480 AlaSerProGluArgPheLeuThrIleGlyAlaAspGlyValGluSerLysProIle 499
 Db 15343 GGATCCACAGAAACGTTTATCATATGAGCCGCAAAATATGAGTCCACAGACGAT 15402
 QY 500 LysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeu 519
 Db 15403 AAGGGGACCGGCGACGCGAAGAAACGCGCGCATGATCAGATTGCGATGAGTAATG 15462
 QY 520 AlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAsp 539
 Db 15463 AGCACTACTACTAGATAGGAGGAGAAATCTGATGATGAGCTTAAGTCCGCTCAT 15522
 QY 540 LeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThr 559
 Db 15523 CTTAATCAGGTATGTGTCCAGGTAGTGCAGTACTGAAATTTATTAAGTTGAGAGC 15582
 QY 560 TyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSer 579

Db 15583 TTCTCTTGTCATCAATTTGGTTTCGACCATACGCGGTCAATTAATGTGAAGAACGTCA 15642
 QY 580 ThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 599
 Db 15643 TCCATGAGGCTATCCGGGACAGTGTTCAGGAGGATCAATGACCGGCGACGGAAGAG 15702
 QY 600 ArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTrpSerGlyAla 619
 Db 15703 CGCAGCATGAAATTTATGCTTATGAGTCACTCGGCTCGTGCTGTACTTCGAGTCCG 15762
 QY 620 LeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleVal 639
 Db 15763 CTTCGATGGTTATCTTTCAGTGTGAGTGAAGTGAATGAGCATTTGTATAGCACCGGGTA 15822
 QY 640 LeuAlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlu 659
 Db 15823 TTACATCAAGAGTGTCTCAATTCGCAATGGCGGGGCAATGTTCTCATTCGTATCCT 15882
 QY 660 GluGluGluPheThrGluThrValValLysAla 670
 Db 15883 TATGCGAAATTTGAAGAGACATTGTCAAGCG 15915
 RESULT 12
 ID ACF71773 standard; DNA; 2019 BP.
 AC 71773;
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens nucleotide sequence #10240.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; biopesticide; virulence factor; disease model; plague;
 KW whooping cough; gene; ds.
 XX
 OS Photorhabdus luminescens.
 XX
 PN WO200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002WO-1B003040.
 XX
 PR 07-FEB-2001; 2001PR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Tasouit S, Glaeser P, Frangeul L, Kunst F, Danchin A;
 PI Buchleser C;
 XX
 DR WPI, 2003-148459/14.
 XX
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 10240; 1205bp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of P. luminescens
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than P. luminescens and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by P.
 CC luminescens. Cells transformed to express the genes are useful for

recombinant production of the proteins, particularly toxins and antibacterial agents useful as insecticides, bactericides and fungicides. The genes, proteins, vectors containing the genes and ab are also useful therapeutically (to treat microbial infection by bacteria or fungi that are sensitive to P. luminescens-encoded toxins or antibiotics) and as biopesticides. Other uses of the genes and the proteins are as virulence factors and for identifying targets of human diseases for which P. luminescens is a model (particularly plague and whooping cough). This sequence represents one of the isolated P. luminescens genes

XX Sequence 2019 BP; 531 A; 386 C; 522 G; 580 T; 0 U; 0 Other;

Alignment Scores:

Score:	2,57e-90	Length:	2019
Percent Similarity:	1400.50	Matches:	301
Beet Local Similarity:	57.76%	Conservative:	86
Query Match:	44.93%	Mismatches:	246
	39.20%	Indels:	37
	10	Gaps:	6

US-10-089-514-2 (1-686) x ACF71773 (1-2019)

```
QY 22 GIUAIAthGlyGlnProProValValProAspAlaAspTrpSerArgLeuPro 41
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7 GAAGTACCGGCGAGTGTGCGTCCGTAGCACCATTCTTGTGACTTATGAAGATTAACG 66

QY 42 ValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArgGluArg 61
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 67 ATAGAGAAATGATGATCGGAGTCTTTCTCCGGGGCCAGACATCCGGCTGAACAGCT 126

QY 62 AspPheGlyIleSerArgAlaIleThrAspSerGlyLeuProValIleuGlyValCys 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 GATTGGGGGTCTGCTGCAAGTTATTGCACATGCCAAGTCCACTGTTGGGTATTGT 186

QY 82 LeuGlyHisGlnGlyIleAlaGlnLeuPheGlyIleValGlyLeuAlaProGluPro 101
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 CTCGGCATACAGGGGATTAACCTGTCATTTGGGGGCTTGAACATCCCCCAATCCG 246

QY 102 MetHisGlyArgValSerGluValArgHisIleGlyGluAspValPheArgGlyLeuPro 121
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 247 GTCATATGCTATGCGATCGCATACAGAAACACTGGAAAGAGATTGCTCGGTGGCT 306

QY 122 SerProPheThrAlaValArgTyHisSerLeuAlaIleThrAspLeuProAspGluLeu 141
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 307 GAACACTTGAAGTGTGATATCACTCTGTTGTATGACCCGGTTGCCGACATCTCG 366

QY 142 GluProLeuAlaIleTrpSerAspArgGlyValMetGlyLeuArgHisArgGlyPro 161
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 AAATGTACTGCGCTGATGATGCTTAATCATGCAATTGACACACCGAAAGCGCG 426

QY 162 LeuTrpGlyValGlnPheHisProGlySerIleGlySerAspPheGlyArgGluIleMet 181
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 ATCTGGGGGTCAATTTCATCTGAATGATGATTCGAATATGGGCGTGACTATCTC 486

QY 182 AlaAspPheArgSerLeuAlaIleHisIleArgAlaArgArgHis----- 197
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 AGTAATTTGTGGCGATACGAGACAGCTTAATAAGAAATATTAACAGAAATGATCCGT 546

QY 198 -----GlyAlaAspSerProTy 203
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 547 CAGAAATATGTACTGATATCAATGACGCTTATCTTGCAAGTAGGGAAGACAAAGTCTC 606

QY 204 GluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArgArgGlyCys 223
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 607 AACCTGAATATTCGTAATGTAAGAGATGATCGATCTGAATCCCTGTATTATCAACGT 666

QY 224 LeuProGlyGluGlyThrThrPheTrpLeuAspSerSerSerValIleuGluGlyAlaSer 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 TATGCAATGATACGATGCTTTCTGGCTTGATGAAATTCATATGACCCCAATGCC 726

QY 244 ArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyIleuThrTyArgVal 263
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 CCGTCTCATGATGGGCGAGCGGTATGAACAAAGGTGCTGCACTTGAAATACAGTGA 786
```

```
QY 264 AlaAspGlyValIleSerValArgGlySerAspGlyThrThrArgThrArgPro 283
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 787 CAGACTGAAATTTGCGTTTATCGTCTCTGATGGCGAGAGATT---CTTACGGAGAT 843

QY 284 PheAspTrpLeuGluGlnIleuGluArgArgValProValaProGluLeu 303
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 844 TTTTTCGCTGATGCTTCACTGATGTTGTTGGTCAATTTCCACCAAAATGAGT 903

QY 304 PropheGluPheAsnLeuGlyTyValGlyTyLeuGlyTyTyLeuValaGluThr 323
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 CCAATTGTTTAAAGCGGTTGTGTGTTTGAATGATGAATTAAGCGCTGACA 963

QY 324 ThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArg 343
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 ATTTGCGAGAACAAATATGCTTCAACAATCTGATGCCAGCTTAATTTACTCCGCA 1023

QY 344 AlaIleAlaLeuAspHisGlnGlyCys-----CysTyLeuLeuAlaLeuAsp 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 TTTTGTCTTCATCATCAGAGATGAACCTTTATGATGTTTATCACCAAAATGAGG 1083

QY 361 ArgArgGlyHisAspAspGlyValaArgAlaIlePheLeuArgIleThrAlaGluThr 380
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 CAGCCACATCAT-----TGTCAACCGGAATTTTCGTTACCAATACA 1125

QY 381 GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1126 -----AAAATACCGGTGAAGTGAATCTGACTTATTTCTGGCCGTGTGAT 1173

QY 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspGlyAspAlaTyLeuLys 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1174 CAATAATAG-----ATTGTCTGAATATGGCGTGAATAATATATGAT 1218

QY 421 ArgIleAspGluCysLeuLysGluIleArgAsnGlyGlySerTyGluIleCysLeuThr 440
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1219 GGAATCCATTAATCACTCCAAATCACTCAAGATGTGTGATCTTAATGAATCTGTGACG 1278

QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTySerAlaLeuArgAla 460
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1279 AATTAGGCCCAAGATGACTTATTCAGACATCATTTGGACGCTTATTCGGCTATGCTCAT 1338

QY 461 IleSerProValProTyGlyAlaLeuLeuGluPheProGluLeuSerValIleuSerAla 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 GCCAGTCTGTACTTATGCGCTTATCTGCGCTGTGGGAGATTTTCTGTACTGATGCA 1398

QY 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerIleProIleLys 500
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1399 TCACCAAGAAACGTTTTTACGTTATGATGACGCCAGAAATATGAGTCCAGACCGATTAG 1458

QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGlyValArgLeuValaAspLeuAla 520
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1459 GGGACCCCGCCACGGGAAAAACGGCAGCCATGATCCGACTTTGCAATGGAATTGACT 1518

QY 521 GlyArgGlyAspArgAlaGluAsnLeuMetIleValaAspLeuValaArgAsnAspLeu 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1519 CAGTCACTAAGAGATAGGGCAGAAAATCGATGATAGTGAATTTAGTGGCCATGATCTT 1578

QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValaGluThrTy 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1579 AATCAGGTATGTCGCCAGGTAGTGTGATGATCTGAATTTATTAAGTTGAGAGCTTC 1638

QY 561 AlaProValHisGlnLeuValSerThrIleArgGlyValArgLeuArgProGlyThrSerThr 580
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1639 TCTTCTGATCATCAATTTGGTTTCACAAACCGGCTCATTTTGTGAAGAAACGTCATCC 1698

QY 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysIleArg 600
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1699 ATAGAGGCTATCCGGGCAATTTTCCAGAGAGATCAATGACCGGCGACCAAGAACGCC 1758

QY 601 ThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTySerGlyAlaLeu 620
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1759 ACGATGAGATTAATGATGCTTATGAGTCAATGCGGCTGTTGTTTACTCCGCGCTT 1818
```

OY		621	GlyTTPPhPdeAlaleuSerGIyAlaIalaIaspLeuSerIlleValIIlaIxThrIlleValIeu	640
Db		1819	GGATGCTTAATCTTTCAAGTGTCGAAGTGCATTGACATTGTGATACGCCACCGGCTATT	1878
OY		641	AAlaSpGIyGlmaIaGLuPhcGIyValGIyAlaIalIeValSerLeuSerApsGInlu	660
Db		1879	CATCAAGAAGTGTGCGTAAATTCGGCATTTGGGGGGCAGATTGTTGCTCATTCGATCCCTTAT	1938
OY		661	GIUGLuPhcThrGIUTh-ValIallIysAla	670
Db		1939	GGCGAAATTTGAAAGACATCGGTCAAACGC	1968
RESULT 13				
ADBO6077			standard; DNA, 2046 BP.	
XX	ADB06077;			
DT	20-NOV-2003	(first entry)		
XX				
DE	Allotiococcus oclidis antigenic protein encoding DNA SEQ ID NO:17.			
XX				
KW	Allotiococcus oclidis; antigenic protein; immunogenic; immunisation;			
XN	gene therapy; Gram-positive bacterium; infection; gene; ds.			
OS	Allotiococcus oclidis.			
XX				
PN	WO2003048304-A2.			
PD	12-JUN-2003.			
XX				
PF	25-NOV-2002; 2002WO-US036123.			
PR	29-NOV-2001; 2001US-0333777P.			
PR	18-NOV-2002; 2002US-0426742P.			
XX				
PA	(AMHP) WYETH HOLDINGS CORP.			
P1	Fletcher LD, Mcmichael JC, Russell DP, Zagursky RJ,			
XX				
DR	WPI; 2003-505284/47.			
PT	P-PsDB; ADB06078.			
XX				
PS	Claim 7; SEQ ID NO 17; 1019pp; English.			
XX				
CC	The present invention describes an isolated polynucleotide (I) of			
CC	Allotiococcus oclidis genomic DNA, which encodes an antigenic protein.			
CC	Allotiococcus oclidis is a Gram-positive bacterium. Also described: (1)			
CC	an isolated polypeptide that is encoded by the polynucleotide (I); (2) an			
CC	expression vector comprising the novel isolated polynucleotide (I), its			
CC	complement, degenerate variant or fragment; (3) a genetically engineered			
CC	host cell, transfected, transformed or infected with the vector of (2);			
CC	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic			
CC	composition comprising the polypeptide, its complement, biological			
CC	equivalent or fragment, or the polynucleotide that is comprised in the			
CC	expression vector; (6) a pharmaceutical composition comprising the			
CC	polypeptide of (1) and a carrier; (7) a protein chip comprising an array			
CC	of the polypeptides of (1), their biological equivalent or fragment; (8)			
CC	immunising against Allotiococcus oclidis by administering to a host the			
CC	immunogenic composition; (9) detecting and/or identifying Allotiococcus			
CC	oclidis in the biological sample; (10) a kit comprising a container			
CC	containing the novel polynucleotide, its degenerate variant or fragment,			
CC	or the antibody of (4); and (11) producing a polypeptide by culturing the			
CC	genetically engineered host cell under conditions suitable to produce the			
CC	polypeptide from the culture. (I) can be used in gene therapy. The			
CC	polynucleotides, polypeptides, antibodies and compositions of the present			
CC	invention can be used for treating and diagnosing diseases, drug			
CC	screening assays and monitoring of effects during drug clinical trials.			
CC	The polynucleotides are useful for expressing and detecting Allotiococcus			

[illegible]

```

Db      832 GCAGTTGAAGCTTCCAGACTGACATCTTTCTTACTTA-----AAAGCCACCGACCA 885
Qy      296 ArgValProValAlaIleProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeu 315
      886 AAATGCCCTTATGAGCGCGCTTGGCTTTGACTTTACAGTAGGCTATATGGCTATATC 945
Qy      316 GlyTyrGluLeuLysAlaGluThrGlyLysAspProAlaHisArgSerProHisPro 334
      946 GGCTATGAGGTCAAAAAGATACAGTTCAAGTCCACACCAAGCAGCAGTACCTATCCG 1005
Qy      335 ASPAlaIlePheLeuPheAlaAspArgAlaIleAlaLeuAspHisGluGlyCysCys 354
      1006 GAGTCTACTTACTTATGATGTCAGCGCTTGGCTATGACCCACGAGAGAGACCTT 1065
Qy      355 TyrLeuLeuAlaLeuAspArgArgGlyHisAspAspGlyValaArgAlaTyrLeuArgGlu 374
      1066 TACTTCTCTCCAC-----CAGGATGAC-----CAGGATTGATTCAGCG 1107
Qy      375 ThrAlaGluThrLeuThrGlyLeuAlaValaArgAlaProAlaGluProThrProAlaMet 394
      1108 GTCAAGACAGGCTCAAG---CAGGACGTCAAGATTGAGCAAGAGAGG----- 1152
Qy      395 ValPheGlyTyrLeuProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAsp 414
      1153 -----CCAGAGCGCAAGCAAGACTTCCACGACTGGCTTCTGTCAA---GAC 1197
Qy      415 LysAspAlaTyrLeuLysArgIleAspGlyCysLeuLeuGluIleArgAsnGlyGlySer 434
      1198 AAAGCAGCTTATTCAAAGACATTGAAACATTCAGACCTGATCAAGGCTGGGGAAGT 1257
Qy      435 TyrGluIleCysLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeu 454
      1258 TATGAGATTGCTTGACCAACCGCTGATATTGAGGCTGACAGTGTGATGAGCTTCGTTAT 1317
Qy      455 TyrSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGlu 474
      1318 TACAAACACTTGGCGAGGAGAAAGTCCCGCTGACTTGGCTTCTTATGATGATGAT 1377
Qy      475 LeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyVal 494
      1378 TTTCAGATTCTTCTCTTCATGAGCGCTTTTATTAATGTCACAAAGACCGGGTTGG 1437
Qy      495 GluSerLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluLysAspGlyArg 514
      1438 ACCACTAAGCCATCAAGGGGACGTCAGACGGGGACAGATGACCAAGAGACCAAGAC 1497
Qy      515 LeuArgAlaAspLeuAlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAsp 534
      1498 TTGATAGAAAGGCTCCGAGTGAAGAGAAAACCAAGGAGAAAACCTGATGATGTCGAC 1557
Qy      535 LeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeu 554
      1558 CTCTTGCGCAATATCTGGCGCTTTTGTGAATCGGGTCGGTTAGGTACCCAAACCTG 1617
Qy      555 PheGluValGluThrTyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeu 574
      1618 ATGATGTGAGAGACTACCTCACCTCCACAGCTGTTTACACGGTTCCAGCCGGGTC 1677
Qy      575 ArgProGlyTyrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetTr 594
      1678 AAAGAGACCTAGATGTGTGAAGTGTGAAGAAACACCTTCCCGGCGGGTCCATGACT 1737
Qy      595 GlyAlaProLysLysArgThrMetGluIleLeuAspArgLeuGluGlyProArgGly 614
      1738 GGGGCGCTAAAGAAAGACCTGGAATATTGATGACTTAAAGCAGAGTCCCAAGAGG 1797
Qy      615 ValTyrSerGlyValaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleVal 634
      1798 ATCTATTCGGGACTATGGCTTTTAAAGCAACATTCACCTTAACTTCAACATCGTC 1857
Qy      635 IleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleVal 654
      1858 ATCCGACACAGTGTGTGMAAGCAGACAGGCAAGCATGGCTGGGGGGTGGATTTGTC 1917

```

```

Qy      655 SerLeuSerAspGluGluGluPheThrGluThrValIleValAlaArgAlaMetVal 674
      1918 ATGCTATCTGATCTCCGAAAGAGAGTTGATGAAAGTGTTTAAAGCTAAGGCGGCTTG 1977
Qy      675 ThrAlaLeu 677
      1978 TCCGCTTG 1986
Db

```

RESULT 14

ADBI2064_00/c	Fragment Name	Begin	End	LOCUS	ADBI2064	Accession	ADBI2064
WP	ADBI2064_00	1	110000				
WP	ADBI2064_01	100001	210000				
WP	ADBI2064_02	200001	310000				
WP	ADBI2064_03	300001	410000				
WP	ADBI2064_04	400001	510000				
WP	ADBI2064_05	500001	610000				
WP	ADBI2064_06	600001	710000				
WP	ADBI2064_07	700001	810000				
WP	ADBI2064_08	800001	910000				
WP	ADBI2064_09	900001	1010000				
WP	ADBI2064_10	1000001	1110000				
WP	ADBI2064_11	1100001	1210000				
WP	ADBI2064_12	1200001	1310000				
WP	ADBI2064_13	1300001	1410000				
WP	ADBI2064_14	1400001	1510000				
WP	ADBI2064_15	1500001	1610000				
WP	ADBI2064_16	1600001	1710000				
WP	ADBI2064_17	1700001	1754382				
ID	ADBI2064 standard; DNA; 1754382 Bp.						
AC	ADBI2064;						
XX	20-NOV-2003 (first entry)						
DT							
DE	Allotococcus oclitis entire genome sequence SEQ ID NO:6651.						
XX							
KW	Allotococcus oclitis; antigenic protein; immunogenic; immunisation; gene therapy; Gram-positive bacterium; infection; gene; de.						
XX							
OS	Allotococcus oclitis.						
PN	MO2003048304-A2.						
XX							
PD	12-JUN-2003.						
XX							
PF	25-NOV-2002; 2002WO-US036123.						
XX							
PR	29-NOV-2001; 2001US-0333777P.						
XX							
PR	18-NOV-2002; 2002US-0426742P.						
XX							
PA	(AMHP) WYETH HOLDINGS CORP.						
PI	Fletcher ID, McMichael JC, Russell DP, Zagursky RJ;						
XX							
DR	WPI; 2003-505284/47.						
XX							
PT	New Allotococcus oclitis polynucleotides and polypeptides, useful for						
PT	treating and diagnosing diseases, drug screening assays and monitoring of						
PT	effects during drug clinical trials.						
XX							
PS	Example 3; SEQ ID NO 6651; 1019pp; English.						
XX							
CC	The present invention describes an isolated polynucleotide (1) of						
CC	CC Allotococcus oclitis genomic DNA, which encodes an antigenic protein.						
CC	CC Allotococcus oclitis is a Gram-positive bacterium. Also described: (1)						
CC	an isolated polypeptide that is encoded by the polynucleotide (1); (2) an						
CC	expression vector comprising the novel isolated polynucleotide (1), its						
CC	complement, degenerate variant or fragment; (3) a genetically engineered						
CC	host cell, transfected, transformed or infected with the vector of (2);						
CC	(4) an antibody specific for the polypeptide of (1); (5) an immunogenic						

CC composition comprising the polypeptide, its complement, biological
CC equivalent or fragment, or the polynucleotide that is comprised in the
CC expression vector; (6) a pharmaceutical composition comprising the
CC polypeptide of (1) and a carrier; (7) a protein chip comprising an array
CC of the polypeptides of (1), their biological equivalent or fragment; (8)
CC immunising against *Allolococcus* otididis by administering to a host the
CC immunogenic composition; (9) detecting and/or identifying *Allolococcus*
CC otididis in the biological sample; (10) a kit comprising a container
CC containing the novel polynucleotide, its degenerate variant or fragment,
CC or the antibody of (4); and (11) producing a polypeptide by culturing the
CC genetically engineered host cell under conditions suitable to produce the
CC polypeptide from the culture. (1) can be used in gene therapy. The
CC polypeptides, polypeptides, antibodies and compositions of the present
CC invention can be used for treating and diagnosing diseases, drug
CC screening assays and monitoring of effects during drug clinical trials.
CC The polynucleotides are useful for expressing and detecting *Allolococcus*
CC otididis. The present sequence represents the entire genome of
CC *Allolococcus* otididis, which is given in the exemplification of the
CC present invention.

XX
SQ Sequence 1754382 BP; 484756A; 391265C; 387369G; 490992T; 0U; 0Other;

Alignment Scores:

Pred. No.:	1.09e-84	Length:	110000
Score:	1350.00	Matches:	296
Percent Similarity:	58.71%	Conservative:	105
Best Local Similarity:	43.34%	Mismatches:	248
Query Match:	37.78%	Indels:	34
DB:	9	Gaps:	12

US-10-089-514-2 (1-686) x ADB12064_00 (1-110000)

QY 1 MetArgThrIleuLeuIleAspAenTYrAspSerPheThriAsnLeuPheGlyIle 20
DB 7807 ATAAAGCTACTCTATTCGATATGATGATTCTTATTAACCTTAACTTACGACTTATT 7748
QY 21 GlyAlaIleThrGlyGlnProValIleValIleProAsp---AlaAspTrpSerArg 39
DB 7747 GGGAAAGTCAACAGGAAAGAGCCATGGTATATAAAATATACAGATGACCTTACCAAGAA 7688
QY 40 LeuProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArg 59
DB 7687 CTATGGATCTCGACTTGTATATGTCATATTTCACCGGGCCGGTATCCGAGCCGG 7628
QY 60 GluArgAspPheGlyIleSerArgAlaIleThrAspSerGlyLeuProValIleGly 79
DB 7627 GACAAAGACTTCGGCTTGGCCGCAAGTGAATGAAAGCTTGACAAAGCTTATCTTGGC 7568
QY 80 ValCysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
DB 7567 ATTGCTCTGGCCACACAGGCGCATTTATTATTATACGGTGGCAACTTAGTGGGGCTGAT 7508
QY 100 GluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGly 119
DB 7507 ATCCCATGATGATGGCGAGAGTTCGGTCTACCATATATGGGAAAAATATCTTTGACGGC 7448
QY 120 LeuProSerProPheThrAlaValArgTyHisSerLeuAlaIleThrAspLeuProAsp 139
DB 7447 ATAGACCAAGGCTTGAAGTCACACGCTACCACTCATGTTGTGTGAAGC-----AAG 7394
QY 140 GluLeuGluProLeu-----AlaTrpSerAspArgGlyValValMetGlyLeuArg 156
DB 7393 GAATTAGAGAAATTTCAATTCGATGCCAAACCGATGATGGAATTGTCAATGCGCTGTCC 7334
QY 157 HisArgGluAspProLeuTrpGlyValGlnPheHisProGlySerIleGlySerAspPhe 176
DB 7333 CACAAACCAAGCCCATCTATGGGCTCCATTCACCCCGAATCCATTGCCACCCAGTAT 7274
QY 177 GlyArgGluIleValAsnPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArg 196
DB 7273 GGGGAAAAATGATGAAAACTTATAGCCTTGTGCAGAGACTACAT----- 7226
QY 197 HisGlyAlaAspSerProTyrgIleuLeuHisValArgArgValAspValLeuProAspAla 216

DB 7225 -----AACCATCCAGCCTTTATTATGAAAAAGTCCGGGAACCTAGACAGC 7178
QY 217 GluGluValArgArgGlyCysLeuProGlyGlyThrPheThrLeuAspSerSer 236
DB 7177 CAAGACCTTACAGCATCAATTTGGCCAAATATGATGACCAAGTACTTGTGCTCGACTCCAGC 7118
QY 237 SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256
DB 7117 AAGGTGAGCCTGGCTTGTGACGTTTCCATCTTGGCATGGCTGACCAAGCGGGC 7058
QY 257 GluTyrlleuThrTyraArgValAlaAspGlyValSerValArgIleSerAspGlyThr 276
DB 7057 CACACTTAATATGATGATTCAGACCAAAAGATGAAAAACAAATATCCGCTCCGCA 6998
QY 277 ThrThrArgThr---ArgArgProPheAsnTyrlleuGluGluGluArgArg 295
DB 6997 GCAGTTGAAAGCTTCCAGACTGACATCTTCTTACTTA-----AAAGCCAAACGACCA 6944
QY 296 ArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrlleuGlyTyrlleu 315
DB 6943 AATGGCTTATGACGCGGCTTGGCTTGACTTTCAGCTGATGATATGCTATATC 6884
QY 316 GlyTyrlleuLeuGlyAlaGluThrThrGly---AspProAlaHisArgSerProHisPro 334
DB 6883 GGTATGAGGTCAAAAAGATACAGTTCAAGTTCACAAACAGACCAAGTCACTTATCCG 6824
QY 335 AspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCys 354
DB 6823 GATGCTTACTTACTTATTTGTGCGGGCTTGGTCTATGACACACAGAAAGAGCCTT 6764
QY 355 TyrlleuLeuAlaLeuAspArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGlu 374
DB 6763 TACTTCTCTCTAC-----CAGATGAC-----CAGATTGATTCAAGC 6722
QY 375 ThrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrProAlaMet 394
DB 6721 GTCAAAAGACAGGCTCAAG---CAGCGATGCAAGTTCCAGCAAGAGAG----- 6677
QY 395 ValPheGlyIleProGluAlaAlaIleGlyPheGlyProLeuAlaArgAlaArgHisAsp 414
DB 6676 -----CCAAAGAGCAAGCAAGCTTCCACGACTGGCTTCTGCAAA---GAC 6632
QY 415 LysAspAlaTyrlleuLeuArgIleAspGlyCysLeuGlyIleArgAsnGlyIleSer 434
DB 6631 AAAGCAGCTTATATCAAAAGACATTGAAACATCCAAAGCTGATCAAGCTGGGGGAAAGT 6572
QY 435 TyrlleuLeuCysLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeu 454
DB 6571 TATGAAGTTTGCTTGAACCAACCGCTGATATTGAGGTCAAGATTGATGGCTTCTTAT 6512
QY 455 TyrservalLeuArgAlaIleSerProValProTyrglyAlaLeuLeuGluPheProGlu 474
DB 6511 TACAAACACTTGGCCGAGAAAGTCCCGGTCAAGTCTGCTTTTGGCTTTAAGTAT 6452
QY 475 LeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyVal 494
DB 6451 TTTCAAGTCTTCTTCTTCCATGAGCGTATTATTAATGAGACAAAGACCGGGGTGTG 6392
QY 495 GluSerIleProIleGlyGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArg 514
DB 6391 ACCACTTAAGCCCATCAAGAGGAGCGGTGACAGCGGGAGACAGATGACCAAGAACCAAGAC 6332
QY 515 LeuArgAlaAspLeuAlaGlyArgGlyAspArgAlaGluAsnLeuMetIleValAsp 534
DB 6331 TTGATAGAAAGGCTCCGCGTGAAGAGAAACCAAGCGAGAAACCTGATGATTGTCCAC 6272
QY 535 LeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeu 554
DB 6271 CTCTGCGCAATGATCTGGCGGTTTGTGAAATCGGGTGGGTGAGTACCAAACTG 6212
QY 555 PheGluValGluThrTyrlleuProValHisGlnLeuValSerThrIleArgGlyArgLeu 574

```

Db      6211 ATGATGTAGAGACCTTACCTCCACCCCTCCACGCTGTTCACGCGTTTCAGGCCGGCTC 6152
Qy      575 ArgProGlyThrSerThrAlaIaCysValArgAlaIaPheProGlyGlySerMetThr 594
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      6151 AAAAGAACCTAGATGTGGTGAAGCTTGAAGAAAACACCTTCCCGGGCGGGTCCATGACT 6092
Qy      595 GAlaIaProlGlyLysArgThrMetGluIleIleAspArgLeuGluGluIyProArgGly 614
      GGGGGCCCTTAAAAAAGAACCCCTGGAATTTATGATGACTTGAAGACAGTCCACAGAGG 6032
Qy      615 ValIySerGlyAlaIaLeuGlyThrPheAlaIaLeuSerGlyAlaIaAspLeuSerIleVal 634
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      6031 ATCTATTCTGGGACCTATTGCTTTTATGACCAACAAATTCACATTAATACCTTCAACATCGTC 5972
Qy      635 IleArgThrIleValIleuAlaIaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleVal 654
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      5971 ATCCGACACAGTGTGTGTGAGACAGACAGACAGATGCGCTGGGGGGGTGGATGTGC 5912
Qy      655 SerLeuSerAspGluGluGluGluPheThrGluThrValValIleAlaArgAlaMetVal 674
      5911 ATGCTATCTGATCTCTGACAGAGAGGTTTGATGAAGTTGTTTAAAGCTAAAGGGCGCTTG 5852
Qy      675 ThrAlaLeu 677
      ::::::::::
Db      5851 TCCGCTTG 5843

RESULT 15
ADB95077
ID      ADB95077 standard, DNA; 2760 BP.
AC      ADB95077;
XX      04-DEC-2003 (first entry)
XX      A. thaliana gene 55483 #SEQ ID 75.
DE      Plant; herbicide; weed; crop field; growth; development; gene; ds.
KW      Arabidopsis thaliana.
OS      WO2003008440-A2.
XX      30-JAN-2003.
XX      16-JUL-2002; 2002MO-EP007929.
XX      16-JUL-2001; 2001US-0305806P.
XX      20-FEB-2002; 2002US-0358416P.
XX      (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX      Levin UZ, Paton DA, Melver JA, Budziszewski GJ, Zhou Q, Aux GW;
PI      Tossberg J, Weigrich Glover L, Ashby CS, Thomas CR, Madhavan E;
PI      Lewis S, Dunn J, Cates E, Law MD;
XX      WPI; 2003-229557/22.
XX      P-PSDB; ADB95078.
PT      Identifying an herbicidal compound, useful for controlling undesirable
PT      vegetation, comprises combining a polypeptide with a compound to be
PT      tested for the ability to bind to the polypeptide or inhibit the activity
PT      of the polypeptide.
XX      Example 9; SEQ ID NO 75; 273bp; English.
CC      The invention relates to a method for identifying a herbicidal compound.
CC      The method of the invention comprises combining a polypeptide having at
CC      least 90% identical to any one of 48 69-1008 residue amino acid sequences
CC      (designated as PI-P48), given in the specification, with a compound to be
CC      tested for the ability to bind to the polypeptide or inhibit the activity
CC      of the polypeptide, under conditions conducive to binding or inhibiting,
CC      respectively. Also disclosed is a method for killing or inhibiting the
CC      growth or viability of a plant by applying to the plant the herbicidal

```

```

CC      compound identified by the novel method, a chimeric construct comprising
CC      a promoter operatively linked to the nucleic acid molecule, a recombinant
CC      vector comprising the chimeric construct and a host cell comprising the
CC      nucleic acid molecule. The method and polypeptides are useful in
CC      screening assays to identify compounds that interact or inhibit the
CC      polypeptides, thus as potential herbicides to control undesirable
CC      vegetation such as weeds in crop fields. Nucleic acid molecules (odd
CC      numbers between ADB95003 and ADB95097) isolated from Arabidopsis thaliana
CC      comprising nucleotide sequences that encode proteins (even numbers
CC      between ADB95004-ADB95098) are essential for plant growth and
CC      development.
XX      SQ      Sequence 2760 BP; 795 A; 492 C; 631 G; 842 T; 0 U; 0 Other;
XX      Alignment Scores:
XX      Pred. No.:      3,83e-82      Length:      2760
XX      Score:      1288.00      Matches:      309
XX      Percent Similarity:      50.48%      Conservative:      115
XX      Best Local Similarity:      36.79%      Mismatches:      238
XX      Query Match:      36.05%      Indels:      178
XX      DB:      10      Gaps:      22
XX      US-10-089-514-2 (1-686) x ADB95077 (1-2760)
Qy      1 MetArgThrLeuIleAspAsnIyAspSerPheThrNIAsnLeuPheGluIyIle 20
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      253 GTGAGGACTTTGTTGATTTGATATGATATGATTTATGATTTATGATGAGCTCG 312
Qy      21 GlyIuAlaIthrGlyGluProProValValValProAsnAspAlaAspTyrSer----- 38
      313 AGTACTATTAAATGAGTGGCTCTCTGCTTATTCGGAATGAT---GAGTGGACGTGGGAA 369
Db      39 -----ArgLeuProValGluAsp-----PheAspAlaIleValIaSerProGly 53
      370 GAAGCTTACCATTTATTTATGAAAGTGTGCTTTGATATATGATTTATGATGCGCTGGA 429
Qy      54 ProGlySerProAspArgGluIuAspPheGlyIleSerArgAlaIleThrAsp--- 72
      430 CCTGATGCGCCATGTGTCACGCTGATAGAAATATGCTTCGCTTTGCTTGAATGC 489
Db      73 SerGlyLeuProValIleuGlyValIyIeGlyIleGlyIleAlaGluLeuPheGly 92
      490 CGTGATATCCCAATTTCTAGGCGCTGCTGCGCCACAGCAGCTAGTATGTCATGGA 549
Qy      93 GlyThrValGlyLeuAlaProGluProMetNIsgIyAspValSerGluValArgIsthr 112
      550 GCTCATGTGGTCATAGCCCCGGAACCAAGTCATGAGCGTGTGAGGATTTGAACATGAT 609
Qy      113 GlyIuAspValPheArgGlyLeuPro-----SerProPheThrAlaValArg 128
      610 GGGAAACATATTGTTTTCGATATTCATCCCGGAGAAACTCGATTTTAAGTGTGATGA 669
Db      129 TyrHisSerIleuAla-----AlaThrAspLeuProAspGlyLeuGluProLeuAlaTyr 146
      670 TACCAATTCATGATCATAGATGAAGATCATCAACAAAGAACTGTACCAATAGCCTGG 729
Qy      147 Ser----- 147
      730 ACGATTATGATGACACTGGCTCTTTCTGTGAGAAATTCCTGTGTCCTGTGAAATAC 789
Db      147 ----- 147
      790 ACTGGAGGCCACTTGGGAAACGGAATCTGTCATCTCTGTTTCAGAAAGTTAGAAATGCA 849
Qy      148 -----AspAspGlyValValMetGlyLeu 155
      850 AGTCATATGGCTTCCTCCCATGTTATGAGGAAACAAGATGACACATTTCTATGGGATC 909
Qy      156 ArgHisLeuGlyIuProlLeuTyrGlyValGlnPheHisProlGluSerIleIySerAsp 175
      910 ATGCATTTCTTTTCCCATATGTTTACAGTTTACATCCAGAAAGATATGCTACTACC 969
Db      176 PheGlyArgGluIleMetAlaAsnPheArgPheLeuAlaLeuAlaHis----- 191

```

```

Db      970  TATGATGTCAGTATTAAATTCAAGGACCTACCTGAATTATGGAGTCGTC 1029
      192  -----HisArgAlaArgArgHisGlyAlaAspSerProTyrGluLeuHisValArgArg 209
      1030  AAATCTACATCCCTCGCTCGAAGAAACATAAATGACACTGCMAAACATGCAGGTG----- 1083
      210  ValAspValLeuProAspAla-----GluGluValArgArgGlyCysLeuPro 225
      1084  -----CCTGATGCTACTCAATTGCTGAAAGAACTTCTGAACTAGATGTACA 1131
      226  GlyGluGlyThr----- 229
      1132  GGAATGCTTCTAGCTATTTTGGGAACCTTAAGTCTGTTTCTGCGAAGACAATGCT 1191
      229  ----- 229
      1192  GTAGAGCTTGTGATATGTGATTCATCATTCMAAACACATACAAATTTGCTGAG 1251
      229  ----- 229
      1252  TTGAAATGCAAGACATGACGCTTTCGCATTAAGTTGTGAGTAAAGAAATATATTT 1311
      230  -----TherPheTrpLeuAspSerSerVal 238
      1312  ATGGAATCTTTGGCAAGAAATAGAGAAATGATACCTTTTGGCTGATCTTCTTCTAGT 1371
      239  LeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyr 258
      1372  GACAAAGCTPAGACGATTTTCTTCATGCGCGGTAAAGTGATCTCTCGAAGCAA 1431
      259  LeuThrTyrArgValAlaAsp-----GlyValValSer 269
      1432  TTGACATTTTGAATTATCTGATCAAAAGTAGGTTACATCAAAACATGCGGACATCTTCTG 1491
      270  ValArgGlySerAspGlyThrThrThrArg-----ThrArgArgProPheAsnTyr 287
      1492  ATGGAAGATTTCTGAGATTTCTACTGAGAAACAATTTCTGAAAGAGCTTTCTTGATTT 1551
      288  LeuGluGluGlnLeuGluArgArgValProValAla-----ProGluLeuProPhe 305
      1552  CTCGGAAGAGGCTTTCATCTATCTATGATGAGAAGACTTCGAAGAGTTGCCCTTT 1611
      306  GluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuValAlaGluThrThrGly 325
      1612  GATTTTGGCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
      326  AspPro-----AlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArg 343
      1669  ATGCCAATTAAATGCTCAAAATCCAAACGCTCCAGATGCAATGCTTTCTTTTGGCGAATAT 1728
      344  AlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeuAlaLeuAspArgArgGly 363
      1729  GTTGCGCATTTGATCATCAACTCGATGACGTTTATATATATTCGCTTACGAAGAGGA 1788
      364  HisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGlu-----ThrLeuThr 380
      1789  -----ACTGCAGAAACCTTCTCCGAATGATACGAAGAGAGCTCATTAAGCTTGATG 1842
      381  GlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGlu 400
      1843  GGTTTTCCCAAGAAAGTTGGAGATCAAACT--CTTCCAGTTATAGATTCACTCTCA 1899
      401  AlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspIysAspAlaTyrLeuLys 420
      1900  TCCAAAACAAGTTTGTCTCT-----GACAAATCCGAGAGCAGTATATCAAC 1947
      421  ArgIleAspGlyCysLeuLysGluIleArgAsnGlyGlySerTyrGluIleCysLeuThr 440
      1948  GATGTTCAAGGCTGATATGATATCAAGAGCGGAGAGCTTGTCTTCTCACT 2007
      441  AsnMetValThrAlaProThr--GluAlaThrAlaLeuProLeuTyrSerAlaLeuArg 459

```

```

Db      2008  ACTCAAAACAGAAAGAAATAGGAATGCTGATCTTTGGAGACTTATCTCACCTGAGA 2067
      460  AlaIleSerProValProTyrGlyAlaLeuLeuGluPhe-----ProGluLeuSerVal 477
      2068  GAGAGGAATCCACACACATATGACGATTTCTCAACTTCTCAAAATCCAAATCTGCTTTTA 2127
      478  LeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLys 497
      2128  TGCTCTTCCCTCCCTAAAGGTTTCTTAAGCTGGACAGAAATGGAATGCTTGAAGCAAA 2187
      498  ProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAla 517
      2188  CCGATTAAAGGTACTATACCTGCTGCTCCACGCTCGAAGAGATGAATTTCTTAATTTG 2247
      518  AspLeuAlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAspLeuValArg 537
      2248  CAATTGAAATCAGTAGAAGAAATCAAGCCGAGAAATCTGATGATGTTGACTTCTTAAGG 2307
      538  AsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluVal 557
      2308  AATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2367
      558  GluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGly 577
      2368  GATCATACACAACAGTATACATACATGATGAGACGATCCGTGACTGAAAAAAACAGAT 2427
      578  ThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaPro 597
      2428  ATTAGTCCAGTGAATGTGTAAGAGCTGCTTCCCTGCGGTTCAATGACTGGTGCCCA 2487
      598  LysLysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSer 617
      2488  AAATTAAGATCTGTGAGATTTCTGATCTCTAGAGAACTGTTCCGAGAGGCTTTTACTCT 2547
      618  GlyAlaLeuGlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgTyr 637
      2548  GGTCAATCGGATATTTCTGATATATGATGATGATGATGATGATGATGATGATGATGATGAT 2607
      638  IleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSer 657
      2608  GTAATTAATCAATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2667
      658  AspGlnGluGluGluPheThrGluThrValValValValAlaArgAlaMetValThrAlaLeu 677
      2668  AGTCCAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2727

```

Search completed: October 5, 2005, 10:48:02
 Job time : 1745.67 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 10:07:08 ; Search time 375.417 Seconds
(without alignments)
2989.973 Million cell updates/sec

Title: US-10-089-514-2
Perfect score: 3573
Sequence: 1 MRTLIDNDSFTNLFXYI.....VVKARAWTALDGSAAVAGAR 686

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapext 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 1202784 seqs, 818138359 residues
Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool_h/US10089514/runat_04102005_105746_8071/app_query.fasta_1.1621
-Database Patents_NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514@cgn2_1_116 @runat_04102005_105746_8071 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPELCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents_NA.*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/prodata/1/ina/backfile1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1816	50.8	2220	3	US-08-765-907A-14
2	1816	50.8	2220	4	US-09-987-614A-14
3	1096.5	30.7	4496	4	US-08-765-907A-6
4	1096.5	30.7	4496	4	US-09-987-614A-6
5	725.5	20.3	2888	3	US-08-765-907A-1
6	725.5	20.3	2888	4	US-09-987-614A-1
7	623.5	17.5	7760	4	US-09-987-614A-1
8	615	17.2	1308	4	US-09-902-540-845
9	615	17.2	1431	4	US-09-902-540-8203
10	578.5	16.2	1101	4	US-09-252-991A-13733
11	570.5	16.0	4403765	3	US-09-252-991A-13413
12	570.5	16.0	4411529	3	US-09-103-840A-2
					Sequence 1, Appli

13	563	15.8	1506	4	US-09-252-991A-811	Sequence 811, App
14	558	15.6	1503	4	US-09-328-352-1525	Sequence 1525, Ap
15	556	15.6	645	2	US-08-403-852D-9	Sequence 9, Appl1
16	556	15.6	645	3	US-08-510-646B-9	Sequence 9, Appl1
17	556	15.6	645	3	US-09-231-818-9	Sequence 9, Appl1
18	556	15.6	645	4	US-09-635-359B-9	Sequence 9, Appl1
19	550	15.4	1518	4	US-09-489-039A-6879	Sequence 6879, Ap
20	548.5	15.4	816	4	US-09-252-991A-13943	Sequence 13943, A
21	510	14.3	2079	4	US-09-248-796A-1387	Sequence 3287, Ap
22	509	14.2	1428	4	US-09-543-681A-3936	Sequence 3936, Ap
23	494.5	13.8	1392	4	US-09-328-352-1856	Sequence 1856, Ap
24	494	13.8	54484	4	US-09-302-540-1272	Sequence 1272, Ap
25	487.5	13.6	1542	4	US-09-540-236-53	Sequence 53, Appl
26	487.5	13.6	49617	4	US-09-596-002-28	Sequence 28, Appl
27	485.5	13.6	1473	4	US-09-902-540-5760	Sequence 5760, Ap
28	474	13.3	1664976	4	US-08-916-421B-1	Sequence 1, Appl1
29	474	13.3	1664976	4	US-09-692-570-1	Sequence 1, Appl1
30	468	13.1	10357	3	US-08-961-527-191	Sequence 191, App
31	465	13.0	1464	4	US-09-583-110-373	Sequence 373, App
32	465	13.0	1464	4	US-09-107-433-751	Sequence 751, App
33	457.5	12.8	2684	4	US-09-252-991A-12457	Sequence 12457, A
34	449.5	12.6	1722	4	US-09-583-110-2581	Sequence 2581, Ap
35	449.5	12.6	1725	4	US-09-107-433-1608	Sequence 1608, Ap
36	449.5	12.6	13926	3	US-08-961-527-5	Sequence 5, Appl1
37	439	12.3	3065	3	US-09-199-637A-129	Sequence 129, App
38	438.5	12.3	1734	3	US-09-486-382B-12	Sequence 12, Appl
39	437.5	12.2	1734	3	US-09-486-382B-10	Sequence 10, Appl1
40	434.5	12.2	1821	3	US-09-486-382B-10	Sequence 10, Appl1
41	433	12.1	1650	2	US-09-001-826-24	Sequence 24, Appl
42	433	12.1	1650	4	US-09-264-854-24	Sequence 24, Appl
43	430	12.0	2161	2	US-09-001-826-4	Sequence 4, Appl1
44	430	12.0	2161	2	US-09-001-826-25	Sequence 25, Appl
45	430	12.0	2161	4	US-09-264-854-4	Sequence 4, Appl1

ALIGNMENTS

RESULT 1
US-08-765-907A-14
Sequence 14, Application US/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAWAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
FILE REFERENCE: Streptogramins
CURRENT APPLICATION NUMBER: US/08765,907A
CURRENT FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 2220
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-14

Alignment Scores:

Pred. No.: 3,336-177
Score: 1816.00
Percent Similarity: 64.03%
Best Local Similarity: 53.47%
Query Match: 50.83%
DB: 3
Length: 2220
Matches: 385
Conservative: 76
Mismatch: 207
Indels: 52
Gaps: 11

US-10-089-514-2 (1-686) x US-08-765-907A-14 (1-2220)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 70 GTGCGAACCTGTGATCGAACATACGACTGTTACCTTACCAACTCTTCCAGATGCTG 129
QY 21 GlyGluAlaThrGlyGlnProValValProAsnAsp-----AlaAspTyrSer 38
Db 130 GCCGAGGTGAACGCGCCCGCTCGTCTGTCGCAACGACGACACCCGACCTGGCAG 189
QY 39 ArgLeuProValAluIleAspPheAspAlaIleValValSerProGlyTyrProGlySerProAsp 58
Db 190 GCCCTGACCGCGGCGCACTTCGACCAACGTCGTGCTTCACCGCGCGCGCACTCCCGCC 249
QY 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78
Db 250 ACCGACACCGAAGCTGGGCTCTGACCAACGTCGTGCTTCACCGCGCGCGCACTCCCGCC 309
QY 79 GlyValCysLeuGlyIleGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98
Db 310 GGGGTGTGCTGGGCGCACGAGCCCTGTGCTGTCGCGCGCGCGCGCGCTGTCACGCA 369
QY 99 ProGluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArg 118
Db 370 CCGGAACCTTTACGCGCGCACGACGACATCCGCGCACGCGGCGCGCGCTGTCGCG 429
QY 119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuPro 138
Db 430 AACATCCCTCCCGCGCGTACCGGTGTCGCTGACCTCGCTGACCGTCCGCGCACTGCC 489
QY 139 AspGluLeuGluProLeuAlaTyrSerAspAspGlyValValMetGlyLeuArgHisArg 158
Db 490 GCCGACCTGGCGCGCACCGCGCGCACCGCGCGAGCTGATGCGCGCTGCGCGCGCACCG 549
QY 159 GluIleProLeuTyrGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178
Db 550 CACCTGCGCGCTTCGCGCGTGCAGTTCACCCGAATCATCGACGGAACACGCGCGCAC 609
QY 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisArg----- 193
Db 610 CGGATGCTCGCGCACTTCGCGCGACTGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCG 669
QY 194 AlaArgArgHisGlyAlaAspSerPro----- 202
Db 670 ACCGACGATACCCGCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 729
QY 203 -----TyrGluLeuHisValArgArgValAspValLeuProAsp 215
Db 730 GCGTCCGCGCGGTGGGAGATACCGCTGCGATGTGCGGAGGTGCGCTGCGCGCGCGAC 789
QY 216 AlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTyrLeuAspSer 235
Db 790 GCGGACCGCGCTTCAACCGCTGTTCGCGACCGCGCGCGCGCGCGGTTCGCTCGACAG 849
QY 236 SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu 255
Db 850 AGCGCGGTGAGCGCGCGCGCTGCGCGCTTCACTTCGCGCGCGCGCGCGCGCGCGCG 909
QY 256 AlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspGly 275
Db 910 GCGGAACATATCATCGACGTCGCGCACCGCGCGCGGTGCGCGTCAAG-----GACG 963
QY 276 ThrThrThrArgThrArgArgPro-----PhePheAsnTyrLeuGluGluGlnLeuGlu 293
Db 964 TCGAGGCGGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1023
QY 294 ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGly 313
Db 1024 GCCCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080
QY 314 TyrLeuGlyTyrGluLeuLysAlaGluThrThrGlyAspProAlaHisArgSerProHis 333
Db 1081 TACCTTGGGTACGAGACCAAGCGCGAGCGCGCGAGGACCGCCACCGCGCGCGAAGTGC 1140

QY 334 ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluCys 353
Db 1141 CCGACGCGGCGCTTATGTTCCGCCGAGCGGAGTCTGCGCTCGACACGACAGAGGCGCG 1200
QY 354 CysTyrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyValArgAla 370
Db 1201 GCGTGTGCTGTGGCACTGAGCAGCACCCGACGCGCGCGCGCGCGCGCGCGCGCGCG 1260
QY 371 TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385
Db 1261 TGGCTACCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1320
QY 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaIleArgPhe 405
Db 1321 CTGCGCGAGACCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1350
QY 406 GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIleAspGluCys 425
Db 1351 -----GTCACACTACCGCGCACGCTGCGCGCGCTACCGGGAAGTGTGAGGAATGC 1401
QY 426 LeuLysGluIleArgAspGlyGluSerTyrGluIleCysLeuThrAspMetValThrAla 445
Db 1402 CCGCGCTGATACCGACGCGCGAGACTACGAGGTGTGCTTACGACAACTGCTCGCGGTG 1461
QY 446 ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro 465
Db 1462 CCGCGCGGATGACCGCGCTTACCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1521
QY 466 TyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAlaSerProGluArgPhe 485
Db 1522 TACGCGCTTACCTGATGATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1581
QY 486 LeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLysGlyThrArgProArg 505
Db 1582 CTGCGCATGCGCGGAGCGGTGTGGCGAGTCAAACTATGAGGACCGCGCGCGCGCGCG 1641
QY 506 GlyGlyThrAlaGluGluAspGlyLysArgLeuArgAlaAspLeuAlaGlyArgGlyLysAsp 525
Db 1642 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1701
QY 526 ArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla 545
Db 1702 CCGAGCGAACACTGATGATCGTTCGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1761
QY 546 IleGlySerValHisValProArgLeuPheGlyValGluThrTyrAlaProValHisGln 565
Db 1762 ATCGGCTCGTCCAGTACCGGCGCGCTGTGAGGTGAGAACCTTACCGCACCGCGCGCG 1821
QY 566 LeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArg 585
Db 1822 CTGCTGACGACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1881
QY 586 AlaAlaPheProGlyIleSerMetThrGlyAlaProLysLysArgThrMetGluIleIle 605
Db 1882 GCGCGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1941
QY 606 AspArgLeuGluGluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeu 625
Db 1942 GACCGCTCGAAGAGGCGCGCGCGCGCGGTGATCGGCGCGCTGCGCTGATCTTCCCTTC 2001
QY 626 SerGlyAlaAlaAspLeuSerIleValIleArgThrIleValIleAlaAspGlyGlnAla 645
Db 2002 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2061
QY 646 GluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGluGluPheThrGlu 665
Db 2062 ACCATCGAGGTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2121
QY 666 ThrValValLysValArgAlaMetValThrAlaLeuAspGlySerAlaValAlaGlyAla 685
Db 2122 ATGCTTCTCAAGCGCGACCAACCTTCCGCGCGCGCG-----CGCGAGGACACCGCGCGCGCG 2178


```

RESULT 2
US-09-987-614A-14
; Sequence 14, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIAUD, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 14
; LENGTH: 2220
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-14

Alignment Scores:
Pred. No.: 3,33e-177 Length: 2220
Score: 1816.00 Matches: 385
Percent Similarity: 64.03% Conservative: 76
Best Local Similarity: 53.47% Mismatches: 207
Query Match: 50.83% Indels: 52
DB: Gaps: 11

US-10-089-514-2 (1-686) x US-09-987-614A-14 (1-2220)

OY 1 MetATGThTleuLeuileAspaantYrraAsPserPhetThiAsenleuPhegniYrile 20
DB :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
OY 70 GTCGCAACCTCTGATTCGACAACTACGACTCGTCACTCAACCTCTTCAGATGCTG 129
DB GTCGCAACCTCTGATTCGACAACTACGACTCGTCACTCAACCTCTTCAGATGCTG 129
OY 21 GtjGluAlaThGjYcInPrProVa1Va1Va1ProAsnsp-----AlaAspTPrSer 38
DB 130 GCGGAGGTGAAGCGCGCGCTCGCTCGTCTGCGACAGACACACCGGACCTGGCAG 189
OY 39 ArgLeuProVa1GluAspPheAspAla1ValSerProGjYrProGjYrSerProAsp 58
DB 190 GCCCTGGCGCGCGGCGACTTCGACAACTGCTGCTCACCGCGCGCGCGCACCGCGC 249
OY 59 ArgGluAArgAspPheGjYrLeSerAArgAla1leThAsPserGjYrLeuProVa1leu 78
DB 250 ACCGACACGACCTGGGCTCGACGCGCGCGCGGTGATACCGAATGGAGCTGCGCTGCTC 309
OY 79 GtjYVa1CyLeuGjYhIsgInGjYl1eAlaGlnLeuPheGjYcGjYrThVa1GjYleuAla 98
DB 310 GGGGTGTGCTGGGCGCACAGGCGCTGTGCTGCTGGCGCGCGCGCGCTGTCTTCACGCA 359
OY 99 ProGjYrProMetHieGjYrArgVa1SerGjYrVa1ArgHieThrGjYcGluAspVa1PheArg 118
DB 370 CCGGAACCTTTCACGCGCGCACAGCGACATCGCGCAACGAGGCGGCGCTGTTGCGC 429
OY 119 GtjYleuProSerProPheThrAlaVa1AgTyrHieSerLeuAla1AlaThrAspLeuPro 138
DB 430 AACATCCCTCTCCCGCTGACCGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 489
OY 139 AspGtjYleuGjYrProLeuAla1TrpSerAspArgGjYrVa1ValMetGjYrLeuAArgHieArg 158
DB 430 GCGGACCTTGGCGCGCACCGCGCACACCGCGCGAGCGGAGCTATGGCGCGCTGCGCGCACGC 549
OY 159 GtjYrVrProLeuTrpGjYrVa1GlnPheHieProGjYrSer1leGjYrSerAspPheGjYrArg 178

```

Db	550	CACCTGCCCCGCTTCGGGTGACAGTTCACCCCAATCGATGACAGGAAACACGGCCAC	609
Oy	179	GLuIleMeTAlAaenPheArGAspLeuAlaLeu-----AlaHisHisArg-----	193
Db	610	CGGATGCTCGGCMACTTCGGGACCTGTCTCTCGGGGGGGGGCCGACACGGCCCCCGAC	669
Oy	194	AlaArGArGHisGlyAlaAspSerPro-----	202
Db	670	ACCGAAGCATACCGGACCCGACCCGGCCCCGGCCCCCGGACCGGACCGGCC	729
Oy	203	-----TyrGluLeuHisValArgArgValAspValLeuProAsp	215
Db	730	GCCTCCGGCCGGATGGGGAGTACCGGTGATGTGCAGGATGCTCGGTGGTCCGAC	789
Oy	216	AlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSer	235
Db	790	GGGAGCGCGCGCTTCACCGCCCTGTTCGCGACGCCCCGGCCGGTTTGGCTTCAGAC	849
Oy	236	SeSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeu	255
Db	850	AGCGCGGTGAGCGGGGCTCGCGCGGTTCACCTTCCTCGGCGCCCCGGCGGCGCTC	909
Oy	256	AlaGluTrpLeuThrTrpArgValAlaAspGlyValSerValArgGlySerAspGly	275
Db	910	GGCAAAACGATCACCCTACGAGCTGCGGACCGGGCCGTGCGCGTCAAG-----GACCGT	963
Oy	276	ThrThrThrArgThrArgArgPro-----PhePheAnTrpLeuGluGluLeuGlu	293
Db	964	TCAAGCGCGGACACCCCGCGGCGCGGACCTCTTCGACCACTTGAAACACAACTGGCC	1023
Oy	294	ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGly	313
Db	1024	GCCGGCGGCGCTGCCC-----GCCACGGGCGTGCCTTCGAGTTCAACTCGGCTACGTGGC	1088
Oy	314	TyrLeuGlyTyrGluLeuLysAlaGluThrThrGlyAspProAlaHisArgSerProHis	333
Db	1081	TACCTCGGCTACGAGACCAAGCGGACGCGCGCGGACGAGCGCCACCGGGGAACTG	1140
Oy	334	ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGluGluGlyCys	353
Db	1141	CCCAACGGCGGCTCATATGTTGGCGACCGAGTGTCTGGCTTCGACCAAGAAAGGGGGG	1200
Oy	354	CysTrpLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyAlaArgAla	370
Db	1201	GCTGGGCTCTCGGACACTGAGACGACCCGAGGCGCGGACCGGACCGCGCGGCAAGC	1266
Oy	371	TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg-----	385
Db	1261	TGGCTCACCGACGCGCGCGGACCTCGGCACACCGCCCCCGCGCGCTTCACCGCTG	1320
Oy	386	AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGlyPhe	405
Db	1321	CTGGCCGACGCAACTGGCCCGCTCGGAC-----	1350
Oy	406	GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIleAspGlyCys	425
Db	1351	-----GTCACCTACCCGACAGCGTGGCGGCTACCGGAACTGCTGAGAAATGC	1401
Oy	426	LeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeuThrAsnMetValThrAla	445
Db	1402	CGCGGCTGATGACCGGACGCGGACACTGACGAGGTGTGCTGAGAAACATCTCCGGGTG	1461
Oy	446	ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro	465
Db	1462	CCCGGCGGATCGACCGGCTACCGGCTACCGGCGCGCTCGGACCGGTACGCCCCCGCC	1521
Oy	466	TyrGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAlaSerProGluAlaPhe	485
Db	1522	TACCGCGCTTACCTGCAGTTCCCGGGGACACCTGTCTCAGCTCTCAACCGAAGGTTTC	1581
Oy	486	LeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLysGlyThrArgProArg	505

Db 1582 CTGCGATCGCGCGGACGGTTGGCGGAGTCCAAACCCATCAAGGACCGCCGCCCG 1641
Qy 506 G1YGIYThrAlaGluIuAspGluArgLeuArgAlaAspLeuAlaGluArgGluIuAsp 525
Db 1642 GGGGGCGGGCCCGCCAGACGGCGGCTCAAGGCTCCCTCCCGCGCGGAGAAAGAC 1701
Qy 526 ArgAlGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAla 545
Db 1702 CGGAGGAGAACTGTATGATGTCGACCTGGTCCGCAACGACTCGGGCCAGCTCTGCGAC 1761
Qy 546 IleG1SerValHisValProArgLeuPheGluValGluThrThrAlaProValHisGln 555
Db 1762 ATGGCTCGGTCCACGACCGGGCTGTTCGAGGTGAGACCTAACCGCACCTGCCACAG 1821
Qy 566 LeuValSerThrIleArgIleArgLeuArgProGluYThrSerThrAlaAlaCysValArg 585
Db 1822 CTGTGACGACGGTCCGGCGCGCTGGCGCGGACGTCCTCCGCGCGCGGGGTACGG 1881
Qy 586 AlaAlaPheProGluIleGlySerMetThrGlyAlaProIleValArgThrMetGluIle 605
Db 1882 GCCGCTTCCCGCGCGGTGATGACCGCGCGCGCCCAAGGTCCGACACATGATTCATC 1941
Qy 606 AspArgLeuGluIuGluIuProArgIValTyrSerG1ValaLeuGluYTrpPheAlaLeu 625
Db 1942 GACCGGCTCGAGAGGGCGCGCGGTGTACTCGGGCGCTGAGGCTTTCGCGCTTC 2001
Qy 626 SerG1ValaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGluGlnAla 645
Db 2002 AGCGGCGCGCGCACTCAGCATGTCATCCGACCATGCTCGCCACCGAGAGGCGCGC 2061
Qy 646 GluPheG1ValaGluIuAlaIleValSerLeuSerAspGluGluIuGluIuPheThrGlu 665
Db 2062 ACCATGGCGGTGGCGCGCGCTGTCGCTGCTCCGACCGCAGCAGGAGTCCCGCA 2121
Qy 666 ThrValValIleValArgAlaMetValThrAlaLeuAspGlySerAlaValaGluAla 685
Db 2122 ATGCTCTCAAGGCGAGACCACTTCGCGCTTG---CGCAGGACACAGCGGGCGCC 2178

RESULT 3
US-08-765-907A-6
; Sequence 6, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-6

Alignment Scores:
Pred. No.: 2,33e-102 Length: 4496
Score: 1096.50 Matches: 236
Percent Similarity: 57.91% Conservative: 57
Best Local Similarity: 46.64% Mismatches: 162
Query Match: 30.69% Indels: 51
DB: 3 Gaps: 10

US-10-089-514-2 (1-686) x US-08-765-907A-6 (1-4496)
Qy 1 MetArgThrLeuLeuIleAspAsnYrAspSerPheThrHisAsnLeuPheGluIuTyrIle 20
Db 3027 GTGCGAACCTTGTGTGATGACAACTACGACTGTGTTCCTTACATCACTTCCAGATGCTG 3086
Qy 21 G1YValaThrG1YGlnProProValValaIleProAsnAsp-----AlaAspTrpSer 38
Db 3087 GCCGAGGTGAAGGGCGCGCTCGGCTGTCGTCCGGAAGAGACACCGCACCTGGCAG 3146
Qy 39 ArgLeuProValGluAspPheAspAlaIleValIleValSerProGluYProGlySerProAsp 58
Db 3147 GCCCTGGCGCGGCGCACTTCCAGAACGTCGTCTCAACCGCGCGCGCGCACCCCGCC 3206
Qy 59 ArgGluArgAspPheGluYIleSerArgArgAlaIleThrAspSerG1YLeuProValLeu 78
Db 3207 ACCGACACCGACCTGGGCTTCAAGCGCGGGGTGATTCACGAATGGGACCTGCGCTGCTC 3266
Qy 79 G1YValCysLeuG1YHisGlnGluIleAlaGlnLeuPheGluYThrValG1YLeuAla 98
Db 3267 GGGGTGTGCTGGGCGACAGGCGCTGTCTGTCGCGCGCGCGCGCGCTGTCACAGCA 3326
Qy 99 ProGluProMetHisG1YArgValSerGluValArgHisThrG1YGluAspValPheArg 118
Db 3327 CCCGAACCTTTCACAGCGCGCACAGCAGCATCCGCCACAGACGAGGCGCTGTTCGCG 3386
Qy 119 G1YLeuProSerProPheThrAlaValArgYrHisSerLeuAlaAlaThrAspLeuPro 138
Db 3387 AACATCCCTCCCGGTGACCGTGTCCGTACACTGCTACCTGTAACCGTCCGCGACCTGCC 3446
Qy 139 AspGluLeuGluIuProLeuAlaTrpSerAspAspG1YValaIleMetG1YLeuArgHisArg 158
Db 3447 GCCGACCTCGGCGCACCGCCACACCGCGCAGGAGAGCTATGCGCTGCCACCGC 3506
Qy 159 G1YValProLeuTrpG1YValaGlnPheHisProGluSerIleG1YSerAspPheGluArg 178
Db 3507 CACCTGCGCGCTTGGCGCGTGCATGCCACCCGAATGATCAGACAGCAACAGGCGCAC 3566
Qy 179 GluIleMetAlaAspPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193
Db 3567 CGGATGCTGCGCAACTTCGCGACCTGTCTCTGCGCGCGCGCGCGCGCGCGCGCGCAC 3626
Qy 194 AlaArgArgHisG1YAlaAspSerPro----- 202
Db 3627 ACCGAACGATACCGGACCGCCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCC 3686
Qy 203 -----TyrGluLeuHisValaArgArgValaAspValLeuProAsp 215
Db 3687 GGTCCGCGCGCGGTGGGAGTACCGGCTGATGTGCGGAGGTGCGCTGCGTGCAGC 3746
Qy 216 AlaGluGluValaArgArgGluCysLeuProGluIuGluYThrThrPheTrpLeuAspSer 235
Db 3747 GCGGACCGCGCTTACCGCGCTTCCGCGCACCGCGCGCGCGCGCGCGCTTCTGCTGCAACG 3806
Qy 236 SerSerValLeuGluIuValaSerArgPheSerPheLeuG1YAspAspArgG1YProLeu 255
Db 3807 AGCGGCTGAGCGCGGCGCTCGCGCTTCACTTCTCTGCGCGCGCGCGCGCGCGCGCTC 3866
Qy 256 AlaGluYrLeuThrTyrArgValaIleAspG1YValaIleSerValArgG1YSerAspG1Y 275
Db 3867 GCGGAACAGATCACCTAGACGTCCGACCGGCGCGGTGCGGTCAAG-----GACGGT 3920
Qy 276 ThrThrThrArgThrArgArgPro-----PhePheAsnYrLeuGluGluIuLeuGlu 293
Db 3921 TCAGGCGGAGACCGCGCGCGCGCGCACTTCTTCAACCTGGAACACACGACTGGCC 3980
Qy 294 ArgArgArgValaProValaIleProGluLeuProPheGluPheAsnLeuG1YrTyrValG1Y 313
Db 3981 GCCCGGCGCTGCC---GCCACCGGCGCTGCGCTGAGTTCACTGAGCTGACGTCGGC 4037
Qy 314 TyrLeuGluYrGluLeuYrAlaGluThrThrG1YAspProAlaHisArgSerProHis 333

Dh 4038 TACTCGGCTACGACAGCCAGCCGCGCGGACGAGCCACCGCGGAACTG 4097
Qy 334 ProAapAlaAaPheLeuPheAaAspArgAlaIleAaLeuAspHisGlnGluGlyCys 353
Dh 4098 CCGGACGGCGCTTCAATGTTGCGGACCGGATGCTGCGCTTCGACACGAAACAGGGCGG 4157
Qy 354 CysArgLeuLeuAaLeuAaP-----ArgArgGlyHisAspAspGlyAlaArgAla 370
Dh 4158 GCCCGGCTCTGCGACTGAGCAGCACCAGCGCGCGCCGACCGGACCGCGCGGAAAGC 4217
Qy 371 TrrLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385
Dh 4218 TGCTCTACCGACGCGCGCGCGGACCTTCCGACACCGCGCGCGCGCTTCACTG 4277
Qy 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaAlaGlyPhe 405
Dh 4278 CTGCGGACGACCAACTGCGCGCTGAC----- 4307
Qy 406 GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIleAspGluCys 425
Dh 4308 -----GTCACTACCGCCACAGCTGCGCGCTACCGGAACTGTGAGAAATGC 4358
Qy 426 LeuLysGluIleAaGlnGlySerTyrGluIleCysLeuThrAsnMetValThrAla 445
Dh 4359 CGCGCGCTGATCACCACGCGGAGACCTACGAGGTGTGCTGACGAAATCTCTCGGAGTG 4418
Qy 446 ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro 465
Dh 4419 CCGCGCGGATCGACCGGCTTACCGCTTACCGCGCGCTGCGGACCGTACGCGCGCGCC 4478
Qy 466 TyrGlyAlaLeuLeuGlu 471
Dh 4479 TAGCGCGCTTACCTGCAG 4496

RESULT 4
US-09-987-614A-6
; Sequence 6, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR APPLICATION NUMBER: 2001-11-15
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 4496
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-6

Alignment Scores:
Pred. No.: 2,33e-102 Length: 4496
Score: 1096.50 Matches: 236
Percent Similarity: 57.91% Conservative: 57
Best Local Similarity: 46.64% Mismatches: 162
Query Match: 30.69% Indels: 51
Dh: 4 Gaps: 10

US-10-089-514-2 (1-686) x US-09-987-614A-6 (1-4496)

Qy 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Dh 3027 GTGGAACTCTGCTGATCGAACAATGACATCTGCTTCAACAACCTTTCAGATCTG 3086
Qy 21 GlyGluAlaThrGlyGlnProProValValProAsnAsp-----AlaAspTyrSer 38
Dh 3087 GCCGAGGTGAACGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3146
Qy 39 ArgLeuProValGluAspPheAspAlaIleValValSerProGlyIleProAsp 58
Dh 3147 GCCCTGGCGCGGCGGCTTGAACAAGTGTGCTTCAACCGCGCGCGCGCGCGCGCGCG 3206
Qy 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValIleu 78
Dh 3207 ACCGACACGACCTTGGGCTTCAACCGCGCGGCTGATATCCGAAATGGACCTGCGCTCTC 3266
Qy 79 GlyValCysLeuGlnHisGlnGlyIleAlaGlnLeuPheGlyIleThrValGlyLeuAla 98
Dh 3267 GGGGTGTGCTGGGCGCACAGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3326
Qy 99 ProGluProMetHisGlyValArgValSerGluValAlaArgHisThrGlyGluAspValPheArg 118
Dh 3327 CCGGAACCTTTCACGCGCGCACAGCGACATCGCCACAGCGGCGGCGGCGGCGGCGG 3386
Qy 119 GlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuPro 128
Dh 3387 AACATCTCTCTCCCGCTGACCGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3446
Qy 139 AspGluLeuGluProLeuAlaTyrSerAspAspGlyValValMetGlyLeuAlaArgHisArg 158
Dh 3447 GCCGACCTGGCGCGCACCGCGCACACCGCGGAGCGGAGCTGATGCGCGTCCGCGCAC 3506
Qy 159 GluLysProLeuTrrpGlyValGlnPheHisProGluSerIleGlySerAspPheGlyArg 178
Dh 3507 CACCTGCGCGCTTGGCGCTGCGGAGTTCCACCGGAAATGATGACAGCGAACAACGCGCAC 3566
Qy 179 GluIleMetAlaAsnPheArgAspLeuAlaLeu-----AlaHisHisArg----- 193
Dh 3567 CGAGTGTGCGCAACTTCCGCGACCTGTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 3626
Qy 194 AlaArgArgHisGlyAlaAspSerPro----- 202
Dh 3627 ACCGAACGATATACCGCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3686
Qy 203 -----TyrGluLeuHisValArgArgValAspValLeuProAsp 215
Dh 3687 GCCTCCGCGCGGTGGGAGTACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3746
Qy 216 AlaGluGluValArgArgGlyCysLeuProGlyGluGlyIleThrThrPheTrrpLeuAspSer 235
Dh 3747 GCGGACGCGCGCTTCAACCGCGCTTCCGCGGACCGCGCGCGCGCGCGCGCGCGCG 3806
Qy 236 SerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyValAspAspArgGlyProLeu 255
Dh 3807 AGCGCGGTGACGCGGAGCTCGCGCGCTTCACTTCTGCGCGCGCGCGCGCGCGCGCTC 3866
Qy 256 AlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgIleSerAspGly 275
Dh 3867 GCGGAACATCATCAACGATCGCGGACCGCGGCGCGCGCGCGCGCGCGCGCGCGCG 3920
Qy 276 ThrThrThrArgThrArgArgPro-----PhePheAsnTyrLeuGluGluGluGlu 293
Dh 3921 TCAGGCGGCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3980
Qy 294 ArgArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyrValGly 313
Dh 3981 GCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4037
Qy 314 TyrLeuGlyTyrGluLeuLysAlaGluThrThrGlyAspProAlaHisArgSerProHis 333
Dh 4038 TACTCGGCTACGACGACCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4097

```
QY 334 ProAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCys 353
|||
|||
|||
Db 4098 CCGGACGGCGGCTTCAAGTTCGCGGACCGGATGCTGCTCGACCAAGAGGGGGCGG 4157
|||
|||
|||
QY 354 CysTyrLeuLeuAlaLeuAsp-----ArgArgGlyHisAspAspGlyAlaArgAla 370
|||
|||
|||
Db 4158 GCGTGGCTCTGGCACTGAGCAGACGCCGCGGCCCGCCAGCCGCGCGGCGGAGCGC 4217
|||
|||
|||
QY 371 TrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaValArg----- 385
|||
|||
|||
Db 4218 TGGGTCAACGACGGCGCGGACCGCTTGCCACCAACCGCCCGCGCTTCACCGCTG 4277
|||
|||
|||
QY 386 AlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGlyPhe 405
|||
|||
|||
Db 4278 CTGCGCGACGACCACTGCGCGCGCTGAC----- 4307
|||
|||
|||
QY 406 GlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeuLysArgIleAspGluCys 425
|||
|||
|||
Db 4308 -----GTCCACTACCGCCACAGCGCTGCGCGCTACCGGAACTGTCGAGGAATGC 4358
|||
|||
|||
QY 426 LeuLysGluIleArgAsnGlyGlnSerTyrGluIleCysIleuThrAsnMetValThrAla 445
|||
|||
|||
Db 4359 CGCGCGCTGTATCAACGACGCGGACCTACGAGGTGCTGTAGAACATGCTCGGGGTG 4418
|||
|||
|||
QY 446 ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValPro 465
|||
|||
|||
Db 4419 CCGCGCGCGGATCGACCGCGCTACCGCGCTACCGCGCGCGCGCGCGCGCGCGCGCC 4478
|||
|||
|||
QY 466 TyrGlyAlaLeuLeuGlu 471
|||
|||
|||
Db 4479 TACGCGCGCTACTCGTACG 4496
|||
|||
|||

RESULT 5
US-08-765-907A-1
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DURUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765, 907A
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

Alignment Scores:
Pred. No.: 2,47e-64 Length: 2888
Score: 725.50 Matches: 150
Percent Similarity: 78.70% Conservative: 20
Best Local Similarity: 69.44% Mismatches: 45
Query Match: 20.31% Indels: 1
DB: 3 Gaps: 1

US-10-089-514-2 (1-686) x US-08-765-907A-1 (1-2888)

QY 470 LeuGluPheProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGly 489
|||
|||
|||
Db 1 CTGACGTTCCCGGGGACACCGGTCTCACTCTCACCGCAAGCGTTCTCGCATCGGC 60
|||
|||
|||
```

```
QY 490 AlaAspGlyValGluSerIleuSerProIleLysGlyThrArgProArgGlyThrAla 509
|||
|||
|||
Db 61 GCGGACGGCTGGGCGGAGTCTCAACCCATCAAGGACACCGCGCGCGCGCGCGCGCC 120
|||
|||
|||
QY 510 GluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsn 529
|||
|||
|||
Db 121 GCCAGAGACGGCGCGCTCAAGGCTCTCGCGCGCGCGAGAGAGACCGACGAGAAC 180
|||
|||
|||
QY 530 LeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerVal 549
|||
|||
|||
Db 181 CTGATATGTCACCTGTGTCGCAAGACGACTCGGCGAGAGTGCACATCGGCTCGCTG 240
|||
|||
|||
QY 550 HisValProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThr 569
|||
|||
|||
Db 241 CACGTACCGGCGCTGTTCAGGTGAGACCTTACGCCACCTCCACAGCTCTGTACAGCG 300
|||
|||
|||
QY 570 IleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPhePro 589
|||
|||
|||
Db 301 GTCCGCGCGCGCTGGCGGCGGACGCTCCGCGCGCGCGCGGCGGCGGCGGCGCTTCCC 360
|||
|||
|||
QY 590 GlyGlySerMetThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeuGlu 609
|||
|||
|||
Db 361 GCGGCGGTCAATGACCGCGCGCGCCAGTCCGCAACCATCGAGTTCATCGACCGCTCGAG 420
|||
|||
|||
QY 610 GluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAla 629
|||
|||
|||
Db 421 AAGGCGCGCGCGCGCTGACTGCGCGCGCGCTGCGCTGCTTACCTCGCTTACCGCGCGCG 480
|||
|||
|||
QY 630 AspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyVal 649
|||
|||
|||
Db 481 GACCTCAGCATGCTCATCCGACCATCGTCCGACCGAGAGGCGGCCACCATCGCGGTG 540
|||
|||
|||
QY 650 GlyGlyAlaIleValSerLeuSerArgGlnGluGluPheThrGluThrValValLys 669
|||
|||
|||
Db 541 GCGCGCGCGCTCTCGCTCGCTCGCTCGGACCGCGAGAGAGGTCCGCAAAATGCTCTCAAG 600
|||
|||
|||
QY 670 AlaArgAlaMetValThrAlaLeuAspGlySerAlaValAlaGlyAla 685
|||
|||
|||
Db 601 GCGGAGACCACTCTCGCGCGCTG---CGCGAGGACACGCGGCGCGCC 645
|||
|||
|||

RESULT 6
US-09-987-614A-1
; Sequence 1, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAWAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DURUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987, 614A
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765, 907
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Alignment Scores:
Pred. No.: 2,47e-64 Length: 2888
```

Score:	725.50	Matches:	150
Percent Similarity:	78.70%	Conservative:	20
Best Local Similarity:	69.44%	Mismatches:	45
Query Match:	20.31%	Indels:	1
Gaps:	4	Gaps:	1

US-10-089-514-2 (1-686) x US-09-987-614A-1 (1-2888)

QY 470 LeuGIuPhaeProGIuLeuSerValLeuSerAlaSerProGIuAryPheLeuThr11 61y 489y

Db 1 CTGAGATTCCCGGGGGCAACCGGTCTCACTCTCTACCCGAACGGTTCTCGCATCGGC 60

QY 490 AlaAepGIyGIyValGIuSerIySProlIyGIyThrAryProAryGIyGIyThr1a 509y

Db 61 GCGCAGCGCTGGGCGGAGTCTCAACCCATCAAGGGACACCGCGCCCGCGCGCGGCC 120y

QY 510 GIuGIuAepGIuAryrGIuAryrGIuAryrGIuAryrGIuAryrGIuAryrGIuAryr 528y

Db 121 GCCCAGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180y

QY 530 LeuWeI1eValAepLeuValAryAeAepLeuAenSerValCySAla11eGIySerVal 549y

Db 181 CTGATGATCGTCGACCTGGTCCGCAACGACCTCGGCGAGGTGTGGACATGCGTCCGTC 240y

QY 550 HisValProAryrLeuPheGIuValGIuThrIyValProValHisGIuLeuValSerThr 568y

Db 241 CAGCTACCGGGCGCTGTTCAGAGTGGAGACTACCGCACCGCTCCACCACTGTCGAGCAG 300y

QY 570 IleAryGIyAryrLeuAryrProGIyThrSerThrAlaIaCyValAryAla1aPhePro 589y

Db 301 GTCCGGCGCGCGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCCC 360y

QY 590 GIyGIySerThrIyValAryProIySlyAryrThrMetGIuIleIleAryrLeuGIu 609y

Db 361 GCGGGGTGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420y

QY 610 GIuGIyProAryrGIyValIyTrSerGIyValAenGIyTrPheAlaLeuSerGIyVala 629y

Db 421 AAGGGCGCGCGCGCGCTGTACTCGGGCGCGGTACTCTTCCGCTTACAGGGCGCGGC 480y

QY 630 AspLeuSerI1eVal11eAryrThrI1eVal11eAryrAepGIyGIuAryrGIuAryr 649y

Db 481 GACCTCAGATCGTCTATCCGACACATCGTCCGACCGACGAGAGCGCGCACCATCGCGCTG 540y

QY 650 GIyGIyVal1eVal1eSerLeuSerAepGIuGIuGIuGIuPheThrGIuThrVal1a11y 669y

Db 541 GGGGGGCGCGTCTCGCGCGCTGTCCGACCCGACGACGAGGTCCGCGAAATGCTCTCAAG 600y

QY 670 AlaAryAlaMetValThrAlaLeuAepGIySerAlaValaGIyAla 685y

Db 601 GCGCAGACCAACCTCGCGCGCGCTG---CGCCAGCGACAGCGGGGGCGC 645y

RESULT 7

US-09-502-340-845/C
: Sequence 845, Application US/09902540

Patent No. 6833447

GENERAL INFORMATION:

APPLICANT: Goldaman, Barry S.

APPLICANT: Slater, Steven C.

APPLICANT: Wiegand, Roger C.

FILE REFERENCE: 38-10(15849)B

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

; CURRENT FILING DATE: 2001-07-10

PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10

NUMBER OF SEQ ID NOS: 16825

; SEQ ID NO 845

LENGTH: 7760
TYPE: DNA

ORGANISM: *Myxococcus xanthus*

US-09-902-540-845

Alignment Scores:

Pred. No.:
Score:

Percent Similarity

Best Local Similarity

Query Match:
DB:

5

US-10-089-514-2 (1-686) x US-09-902-540-845 (1-7760)

207 ValArgValAspValLeuProAspAlaGluGluVal

Db	4738	CTGCGTCCCGTGGACTCTG-----	-----CGAGCCCTCCACGC	470
Qy	227	GIuGIYThrThrPheTrIleuAAsrSerSerValIleuGIuIyAlAsrArgPheSer	246	
Db	4702	GAA---AGCGGTTCTCTCTGAGAGCAACCGGGT---AGCGCGAGGGCGCTATTTC	464	
Qy	247	PheIleuGIYAAsrAsrArgGIYProIleuAlaGIuTrIleuThrTrArgValAlaAsrGIY	266	
Db	4648	TTCTCTGGG-----CGAAGCCCTTCTCGCTTTCACCCGGAAAGAGGA	460	
Qy	267	ValValSerValArgGIYSer-----AspGIYThrThrThrArgThrArgArgProPhe	284	
Db	4603	CAGTCTTCATTGATGAGGGCGCTCCAGAGGGTACCGCGACGAGTG-----	455	
Qy	285	PheAsnTrIleuGIuGIuGIuIleuGIuArgArgValProValAlaProGIuIleuPro	304	
Db	4555	-----CTCGCTCGCTGCTGGCGGGGTGGCGTGCCTCACTCCCGAATCCG	450	
Qy	305	PheGIuPheAsnIleuGIYTrValGIYTrIleuGIYTrGIuIleuValAlaGIuThr	324	
Db	4504	CTC---TTTCTGGCGCGCGCGGTGGAGCTTCTCGAGTTCAGAGCCGGCGCACTGTTTGAG	444	
Qy	325	GIYAAsrProAlaNIAsrSerProNIAs-----ProAspAlaAlaPheIleuPheAla	341	
Db	4447	TTCCTGCCCCGACCCACAGGGACAGCTGCACCTGCACCGGAATGCAGCTGTCTATTATC	438	
Qy	342	AsrArgValAlaAlaIleuAsrNIAsGIuGIuIYAsrGIYTrIleuIleuAlaIleuAsrArg	361	
Db	4387	GACACCTTCTACAGCTGAGCACCAACAGGGCCAAATGTCTCTGGGGCACGGCTCG	432	
Qy	362	ArgGIYNIAsrAsrArgGIYAlaArgAlaTrIleuArgGIuThr---AlaGIuThrIleuThr	380	
Db	4327	GACTGGAGAGACTGTGAGCGCAGAGTGGACGCTTGAAGGTCCACGTCCGACCGCGACG	426	
Qy	381	GIYIleuAlaValaArgAlaProAlaGIuProThrProAlaMetValPheGIYIleuProGIu	400	
Db	4267	CCACAGCGCACGCGGTGCCA-----CCCGCATG-----	423	
Qy	401	AlaAlaIaGIYPhGIYProIleuAla-----ArgAlaArgNIAsAsrAlaIaTr	418	
Db	4237	-----GATGGCTCCACCTCGCGCGATAGTGTCCAACTTCACTCGATGGCTTAC	419	
Qy	419	IleuYAsrGIaAsrPGLuCyAsIleuYAsGIuIleArgAsnGIuIleuSerTrYGluIleCyAs	438	
Db	4189	CTGGAGCGCACTGAGACGGGTCCGCGAGACATCCGCCCGGTAGCACTGATCAAGTCAAT	413	
Qy	439	IleuThrAsnMetValThrAlaProThrGIuAlaThrAlaIleuProIeuTrYSerAlaIleu	458	
Db	4129	CTCTGCGACGGCTGAGAGTGAAGACTTTCGCCGTAGAGCGCGCTCGCGCTCTCGAGAGCGTG	407	
Qy	459	ArgAlaIleSerProValProTrYGluValaIleuIleuGIuPheProGIuIleuSerValIleu	478	
Db	4069	TTCGCCATATGACCCGCTCACTTCCAGCATATTCGAAGGTAGAGGCTTCCACGTCTGC	401	
Qy	479	SerAlaSerProGIuArgPheIleuThrIleGIYAlaAspGIYAlaGIuIleuSerTrPro	498	
Db	4009	AGGAGCTCACCGAGACGGCTGATGGATGTG---GAGAAAGCGCGAGCACACGCGGCC	395	
Qy	499	IleIleGIYThrArgProArgGIYGIYThrAlaGIuIleuAspGIuArgIleuArgAlaAsp	518	

```
Dh      3952 ATGCGCGGAAGCGC---CGACGGGGGACCGCCGGAAGAGACGGCGCTTCCTCACGAG 3896
Qy      519  LeuAlaGlyArgGluYrAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
Dh      3895 CTCGGCACCAAGAGAGAGCGCGCGAGACACCGCATGCTGTGGACCTCGAGCGGAC 3836
Qy      539  AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
Dh      3835 GACCTGGCGCGGGGTGTCACCTTAGCTGCTCGAGGTGACGAAGCTGATGAGCATCGTC 3776
Qy      559  ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
Dh      3775 GAGTAGCGCCACGCTCTCCACATTGAATCGAAGGTCTCGGCCAGCTCGCGCGGGGTG 3716
Qy      579  SerThrAlaIaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Dh      3715 GACCCGCTGGACGCGGTGGGGCGGCTCTTCCAGCGCGAGACGATTACGGCGCGCGAG 3656
Qy      599  LysArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGly 618
Dh      3655 ATACGCACGATGCATAATCATCACCGAGCTGAGCCACATGCGCGGGCTCTATACGGGC 3596
Qy      619  AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Dh      3595 TCCCTGGGGTACCTGTCCTTCACGGGTGATTGGATTGTAACATCGTCATCCGACGCTG 3536
Qy      639  ValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValIleValSerLeuSerAsp 658
Dh      3535 GTGTGTAAGACCGCGCGGCTTACGCGCAGGTGGGAGCGGCATCGTCCACGACTCGCAG 3476
Qy      659  GlnGluGlnGluPheThrGluThrValIleValIleValArgAlaMetValThrAlaLeu 677
Dh      3475 CCCCAGACAGAAATACAGAGAGACGCTCAACAAGCGCGCTCGCAGCTCTGACCCCTG 3419

RESULT 8
US-09-902-540-8203
/ Sequence 8203, Application US/09902540
/ Patent No. 6833447
/ GENERAL INFORMATION:
/ APPLICANT: Goldman, Barry S.
/ APPLICANT: Hinkle, Gregory J.
/ APPLICANT: Slater, Steven C.
/ APPLICANT: Wiegand, Roger C.
/ TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
/ FILE REFERENCE: 38-10(15849)B
/ CURRENT APPLICATION NUMBER: US/09/902,540
/ PRIOR FILING DATE: 2001-07-10
/ PRIOR APPLICATION NUMBER: 60/217,883
/ NUMBER OF SEQ ID NOS: 16825
/ SEQ ID NO 8203
/ LENGTH: 1308
/ TYPE: DNA
/ ORGANISM: Myxococcus xanthus
US-09-902-540-8203

Alignment Scores:
Pred. No.:      1,8e-53      Length:      1308
Score:          615.00      Matches:      169
Percent Similarity: 50.77%      Conservative: 62
Best Local Similarity: 37.14%      Mismatches:  186
Query Match:     17.21%      Indels:      38
DB:              4          Gaps:         12

US-10-089-514-2 (1-686) x US-09-902-540-8203 (1-1308)
Qy      231  PheTrpLeuAspSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAsp 250
Dh      10  TTCCTGCTGAGAGACCGGGTC---AGCGCGAGGGGCGCTATTGTTCTCTCGGC--- 63
Qy      251  AspArgGlyProLeuAlaGluTyrLeuThrTyrArgValAlaAspGlyValSerVal 270
Dh      64  -----GCGAAGCCCTTCCTGCTTTCACCGCGAAGAGGAGACAGTCTTCATT 111
```

```
Qy      271  ArgGlySer-----AspGlyThrThrThrArgThrArgProPheAsnTyrLeu 288
Dh      112  GATGGCGCCTCCACAGAGGTACGCGCGAGAGGTG-----CTG 150
Qy      289  GlnGluGlnLeuGluArgArgValProValAlaProGluLeuProPheGluPheAsn 308
Dh      151  CGCTCGCTCTGGCGCGGTGGCGGTGGCGGTCAATCCGGAATGCGGCTC---TTCCTG 207
Qy      309  LeuGlyTyrValGlyTyrLeuGlyTyrGluLeuValGluIleThrGlyAspProAla 328
Dh      208  GCGGCGCGGTGGGCTTCTTCGCTTACGAGCGCGCACTGTTTGAAGTCCCTGCCCCG 267
Qy      329  HisArgSerProHis-----ProAspAlaAlaPheLeuPheAlaAspArgAlaIle 345
Dh      268  CACCCACGGGAGCAGACTGCACCTGCGCCGGAATGCACCTGTCAATTCAACACCTTCG 327
Qy      346  AlaLeuAspHisGlnGluGlyCysTyrTyrLeuAlaLeuAspArgArgGlyHisAsp 365
Dh      328  ACGGTGACCAACACAGAGGCGCAATTGCTCTGTGGCCACGCGGCTCGACTGGAGAGAC 387
Qy      366  AspGlyValArgAlaTrpLeuArgGluThr---AlaGluThrLeuThrGlyLeuAlaVal 384
Dh      388  TGTGACGCAAGTTGAGCGCTTGAAGTTCACGTCCGACCGACCGCACGCGGCCA 447
Qy      385  ArgAlaProAlaGluProThrProAlaMetValPheGlyIleProGluAlaAlaGly 404
Dh      448  CGCGTGCCA-----CCCGGATG----- 465
Qy      405  PheGlyProLeuAla-----ArgAlaArgHisAspLysAspAlaTyrLeuLysArgIle 422
Dh      466  GATGGGCGCCACCTCGCGGTATCGCTCCCACTTCACTCCGTGATTCACGACGCAAGTC 525
Qy      423  AspGluCysLeuLeuGluIleArgAsnGlyGlySerTyrGluIleCysLeuThrAsnMet 442
Dh      526  GAGCGCGTGGCGGAGTACATCCGCGCGGTGACACGTATCAAGTCAATCTTCGACGGG 585
Qy      443  ValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSer 462
Dh      586  CTGAGAGTGAGACTTTCGCGGTGAGCGGCTCGGCTCTTACGAGCGGTGCCGACATCAC 645
Qy      463  ProValProTyrGlyValAlaLeuGluPheProGluLeuSerValLeuSerAlaSerPro 482
Dh      646  CCGGTCACTTCGCGCATCTCGAAGGTGACGCGCTTCACGTCTTCAGCGCTCACC 705
Qy      483  GluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLysGlyThr 502
Dh      706  GAGCGGCTGATGGTGTG---GAGAACGGCCGAGACCAACGAGGCGCCATCGCGGAGAC 762
Qy      503  ArgProArgGlyGlyTyrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArg 522
Dh      763  CGC---CGACGGGGCACGCGCCGAGAGGAGAGCGCGCTTGTCACACGAGCTGGGCACAC 819
Qy      523  GluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgHisAspLeuAsnSer 542
Dh      820  GAGAAAGACGCGCGCGACGACGCGCATGCTGTGACCTTGACGCGCAACCACTGGGCGCG 879
Qy      543  ValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyrAlaPro 562
Dh      880  GTGTGACCTTAGCTGTGCTGAGTGAAGCAAGCTGATGACATCTCGAATGAGCCAC 939
Qy      563  ValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaIa 582
Dh      940  GTCTTCACATTGAATCGAAGTCTGTGGCGCACTCGCCCGGCGCTGAGCCGCTGGAC 999
Qy      583  CysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrMet 602
Dh      1000  GTGTGCGCGCGCTTTCACAGCGGAGCAATTAACGGGCTGCGAAGATGACAGATG 1059
Qy      603  GluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGlyValLeuGlyTyrP 622
Dh      1060  CAATCATCACCGAAGCTGAGCACATGCGCGCGGCTCTTATACGGGCTCGTGGGTAC 1119
```


Qy 623 PheAlaLeuSerGlyAlaAlaPheLeuSerIleValIleArgThrIleValLeuAlaAsp 642
Db 1120 CTGCTCTTCAAGGGGTGATTGATTTGACATCTCTATCCGACGCTGGTGGTGAAGAC 1179
Qy 643 GlyAlaIleAluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGluGlu 662
Db 1180 GGGCAGGCGCTACGCGCAGTGGGAGCGCGCATCGTCCACAGCTCGCAGCCCGCAGAA 1239
Qy 663 PheThrGluThrValValIleValAlaArgAlaMetValThrAlaLeu 677
Db 1240 TACAAGAGAGCTCAACAAGCGCGCTCGACGCTCGTGGCCCTG 1284
RESULT 9
US-09-252-991A-13733
; Sequence 13733, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13733
; LENGTH: 1431
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13733
Alignment Scores:
Pred. No.: 2,08e-53 Length: 1431
Score: 615.00 Matches: 168
Percent Similarity: 48.26% Conservative: 54
Best Local Similarity: 36.52% Mismatches: 184
Query Match: 17.21% Indels: 54
Gaps: 11
US-10-089-514-2 (1-686) x US-09-252-991A-13733 (1-1431)
Qy 233 LeuAspSerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspArg 252
Db 160 CTGATTCGGGTGCGCCCTGATCCCTGACGCTATGACCTTATG-----AGCGCC 213
Qy 253 GlyProLeuAlaGluThrValThrValAlaAspGlyValAlaSerValArgGly 272
Db 214 TGCCCATTCGACAACTGGCGCGCTG----- 240
Qy 273 SerAspGlyThrThrArgThrArgArgProPhePheAsnTyLeuGluGlu----- 290
Db 241 CCGGAGAAATCGGCAACGCG-----TTCTTCGCCCGCGCTACCGCAAGCGCC 288
Qy 291 ---GlnLeuGluArgArgValProValAlaProGluLeuProPheGluPheAsnLeu 309
Db 289 GGAAGCGTGGCGCCCGCGCAGCTGCGCCGCTACCAACTGCTTCGCGCGC----- 342
Qy 310 GlyTyValGlyTyLeuGlyTyGluLeu-----LysAlaGluThrThrGlyAspPro 327
Db 343 GGTCTGCTGGGCTATTATCTACGACTGGTGGCGCATCGAGCGCATCGCGCAACAC 402
Qy 328 AlaIleArgSer---ProHisProAspAlaAlaPheLeuPheAlaAspArgAlaIleAla 346
Db 403 GCCCGCATGACCTGGCGCTCGCGCTGCTCCGTCGCGCTGATGCTGGCGCTGATC 462
Qy 347 LeuAspHisGlnGluGlyCysCysTyLeuLeu-----AlaLeuAspArgGly 363
Db 463 AGGACACACACGCGCGCAGCAGCAACTGTGATTCACCGCGATGAGCAGCGGAA 522
Qy 364 HisAspAspGlyAlaArgAlaTyrPheArgGluThrAlaGluThrLeuThrGlyLeuAla 383

Db 523 CGCCAGCGCTGATTCGCCCTGTTACAGCAGAAACG----- 558
Qy 384 ValArgAlaProAlaGluProThrProAlaMetValPheGlyTyrProGluAlaAla 403
Db 559 -----CAAGCGCGCGCGCG 573
Qy 404 GlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAla-----TyrLeuLysArg 421
Db 574 AACTTAAGCTTGTGGAAATTCGACGAGCATAGCGCATCCACTATTCGCGAGCG 633
Qy 422 IleAspGluCysLeuLysGlnIleArgAsnGlyGluSerTyrcTyrIleCysLeuThrAsn 441
Db 634 ATCCGCGCATCCAGACATCACTCCAGGAGCGGACGTGCTACAGTGAACTATACCG 693
Qy 442 MetValThrAlaProThrGluAlaThrAlaLeuProLeuTySerAlaLeuArgAlaIle 461
Db 694 CGCTTCAGGCGCTTCGACGCGGCTCGCGCTGCGCTATCGCCCTCGCGCGAGCGC 753
Qy 462 SerProValProTyrglyAlaLeuLeuGluPheProGluLeuSerValLeuSerAlaSer 481
Db 754 TGCCCGACGCGCTTCGCGCTACCTGCGACTGGCGCGCGCGCATCTCAGCTGTGC 813
Qy 482 ProGluArgPheLeuThrIleGlyAlaAspGlyValGluSerTyrcTyrIleGly 501
Db 814 CCGAGCGCTTCTCTCAAGCTCAC---AAGGCAAGTGGAAACCGCGCATCAAGGCG 870
Qy 502 ThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuAlaAspLeuAlaGly 521
Db 871 ACCCGCGCGCGCGCAAGACCCCGGAGAGACATGCGCTGGCGCGCTGCTGCGGC 930
Qy 522 ArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsn 541
Db 931 AGCCCAAGACCGCGCGGAAACCTGATGATCTGACCTGCGCAACGACATCGA 990
Qy 542 SerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyrAla 561
Db 991 CGGATTCGCAACTGGCAGCTGACGCGTACCGAGCTTTCGCCGGAACCTATCC 1050
Qy 562 ProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAla 581
Db 1051 AAGTCATCATCTGATGACAGCTGACCGCGCACTGCGCGCGCGCAAGATTCGCGCC 1110
Qy 582 AlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThr 601
Db 1111 GACCTCGTGAAGCGCTTCCCGCGCTGATCAGCGCGCGCGCAAGATTCGCGCC 1170
Qy 602 MetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrcSerGlyAlaLeuGly 621
Db 1171 ATGCAATCATCGACCACTGGAACTGGAACTGCAACCGACCGAGCGGCACTACGCGACGCTGTC 1230
Qy 622 TrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAla 641
Db 1231 TACCTGAGCTGGCGCGGAGATGACAGCTGATGATCGCATCGGACCTGCGTCAAG 1290
Qy 642 AspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGlnGluGlu 661
Db 1291 AACGCGCAGGTGAGTTCGTGGGCGCGCGCGCATGTCGCGACTCGCACTGGAGGAGAC 1350
Qy 662 GluPheThrGluThrValValIleValAlaArgAlaMetValThrAlaLeuAspGlySerAla 681
Db 1351 GAGTACAGAAACCTGGCAAGAGTCCGGGTGCTGCGAAACCTGGAAAGGATGGCC 1410
RESULT 10
US-09-252-991A-13413/c
; Sequence 13413, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196, 136
; CURRENT APPLICATION NUMBER: US/09/252,991A

QY 122 SerPro-----PheThrAlaValArgTyHisSer--- 131
 Db 1798585 AGTCCCGGCTGCATGGGGCAGTCTCGAAGTCCGAGTCCGCTGCATCCGGTTTCAATATGCT 1798644
 QY 132 LeuAlaAlaThrAspLeuProAspGluLeuGluProLeu-AlaTrpSerAspArgLy-- 150
 Db 1798645 CAGGTACACACCGATGGCCGACTCTTGGCCGCGGCTTCCGGCTGGGGGGTGGCCCAT 1798704
 QY 151 -----ValMetGlyLeuArg-HisArgGluysProLeut 163
 Db 1798705 GCGCGCCGTTGGCCGGCGGATGATGCTGCTGGGCTACGCCGGGCGCATGCGATCT 1798764
 QY 163 rPGLyValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIleMetAla 183
 Db 1798765 GGGAAAGATGGTTCCGTCGACCGCCGACCGACGACGACCACTCCGGTAG-----G 1798815
 QY 183 snPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisGlyAlaAspSerProt 203
 Db 1798816 ATTTCGGCTCTGCGCG--GCCAGACACCGGTC----- 1798847
 QY 203 yTGLuLeuHisValArgArgValAspValLeuProAspAlaGlu-----GluValA 220
 Db 1798848 -----GTTCCGGTGACTGCG--AAGGCTTGGCCGACAGAGAGACGCCGCTGGCGCT 1798899
 QY 220 rGArgGlyCyValuLeuProGluGluGlyThrTrpPheTrpLeuAspSerSerValLeuG 240
 Db 1798900 ACCCGACCTCGCCCGCAATCGCCCGGATGTTCTGCTGAAGTCGGCCAGAACCGCC 1798959
 QY 240 IuGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyLeuT 260
 Db 1798960 GGTGTGTCCGAGATGCTGTTTATCGTTCGGCGGG--GCCCAACGGCG-----TTGA 1799010
 QY 260 hrTYArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgT 280
 Db 1799011 CCGTGCGT-----GAGGGGCAAGCGGATGCTGGGTCGGTCCGCCAAGACCTCCCA 1799064
 QY 280 hrArgArgProPhePheHisnTyTLeuGluGluGlnLeu-----A 294
 Db 1799065 CTGGCGGAGACCCCGCTCGGGCGCGCTGACGTGACTTGGACTCTGCGTACCGCGGATC 1799124
 QY 294 rGArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyTValGlyT 314
 Db 1799125 GTCAAGTCGACACCGGGCTTTCGCGCGCTG-----TGGGTGGCATGGTGGGTT 1799172
 QY 314 yTLeuGlyTyTLeuLeuValaGluThrThrGlyAspProAlaHisArgSerProHis- 333
 Db 1799173 TCTTCGCTTATGACATGGTGGAGCG-----CTGGACGATGTCGGAAC 1799217
 QY 334 -----ProAspAlaAlaPheLeuPheAlaAspArgAlaIleA 346
 Db 1799218 GGGCGCTGCATGACTCTGCTCGCGGACATGCTGCTTCTGGCCACCGCATGTGGCGG 1799277
 QY 346 IaleuAspHisIleGlnGlyCyValSerTyTLeuLeu-----AlaLeuAspArgArgLyH 364
 Db 1799278 CCGTGCATACACAGAGGGGACCAATCAACGTGATAGCCGCAACGCGCTGAATCGGAACGCA 1799337
 QY 364 IAspAspGlyValaArgAlaTrpLeuArgGluThrAlaGluThrLeuThrGlyLeuAlaV 384
 Db 1799338 CCGACGAGCGGGTCT--GACTGGGCGTACAGACGACGCGGTGCTCGGCTGCACTGTATGA 1799394
 QY 384 aIArgAlaProAlaGluProThrProAlaMetVal-----PheGlyTLeProGluAlaA 402
 Db 1799395 CCGGACGGCTCGGCAACCACTACCTGCAACCGTGGCCACTTACGCGCACCCGAG--- 1799450
 QY 402 IaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspArgValAspAlaTyTLeuValArgI 422
 Db 1799451 -----CCGGCCACCGTGGCAGCAACGCAACCGTGAAGAAATATGTCGATCG 1799496
 QY 422 IaAspGlyCyValuLeuysGluIleArgAsnGlyGluSerTyTLeuLeuThrAsnM 442
 Db 1799497 TCGAATACCTGTGGATCAATTCGACGCGGTGAAGGTTTCAGTGTGTCCTCGCAGC 1799556
 QY 442 etValThrAlaProThrGluAlaThrAlaLeuProLeuTyTSerAlaLeuArgAlaIleS 462

Db 1799557 GCTTCGAGATGGACACCGAATGTCGATCCCAACGCTGATCCGAATTCCTGGGGTAAACA 1799616
Qy 462 ePProvalProtyGlyAlaLeuLeuGluPhePro-----GluLeuS 476
Db 1799617 ACCCAAGCTCCATCATGTATCTACTGACAGGCGCGAATAGATGATGATGACATCTTT 1799676
Qy 476 eVAlleuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluS 496
Db 1799677 CGATTGTGGATTCAGATCCGAGCGCTGGTAACGGTCT---CACGAAAGCTGGCGACGA 1799733
Qy 496 eLySProIleLyGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuA 516
Db 1799734 CGATCCGATCGCCGAAACCGGTGGCGGAGAGACAGACGAGAGACGTCCTTCTCG 1799793
Qy 516 rGAlaAspLeuAlaGlyArgGluLyAspArgAlaGluAnMetIleValaAspLeuV 536
Db 1799794 AAAAAGACTCTGTCGGCAGACACAAAGACGTCCGAGCATGTATCTGTGCACTTCG 1799853
Qy 536 AlArgaAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheG 556
Db 1799854 GCCGAAACGACTGCTGGGTGTGCACGCGGGCACTGTTCGGGTGAGGATTACAGCC 1799913
Qy 556 lUValaGluThrTyrAlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgP 576
Db 1799914 ACATCGACGGTACAGCACGTGATGACACTGTGTCTACAGTGACCGGAAAGCTCGCG 1799973
Qy 576 roGlyThrSerThrAlaAlaCysValaArgAlaAlaPheProGlyGlySerMetThrGlyA 596
Db 1799974 AAGGGCCGACCGGCTGAGACCGGTACCGCTCTTCCGGCGGACACGCTGTGGGGCG 1800033
Qy 596 lAProLyLyArgTyrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValT 616
Db 1800034 CGCGAAGAGTGGCGGCGCATGAGACTGATGCAAGAGGTGAGGAAGACACGCGCGGCTTT 1800093
Qy 616 yRSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleA 636
Db 1800094 ACGCGGTGTGTCTGCTTACCTTGACTTCGCGGACACGCTGACTTCGCAATCGCCATCC 1800153
Qy 636 rGTThrIleValaLeuAlaAspGlyGluAlaGluPheGlyValGlyAlaIleValSerL 656
Db 1800154 GCACCGCGCTGATGCTGATACGGCACCGCTTATGTCCAGCGAGGGGGTGTGTGGCGG 1800213
Qy 656 eUSeArgPngInGluGluPheThrGluThrValVallyAsAlaArgAlaMetValThra 676
Db 1800214 ACTCCAAAGGATCTACGAATATCAAGACGACGCGAAGCAAGGCTCGGCTGTGTCTACAG 1800273
Qy 676 lAlaAspGlySerAlaValAlaGlyAla 695
Db 1800274 CGATCGCTGCGCGGACGCTGCGGCT 1800302

RESULT 12
US-09-103-840A-1
Sequence 1, Application US/09103840A
Patent No. 6294328
GENERAL INFORMATION
APPLICANT: FLEISCHMAN, Robert D.
APPLICANT: WHITE, Owen R.
APPLICANT: FRASER, Claire M.
APPLICANT: VENTER, John C.
TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
FILE REFERENCE: 24366-20007.00
CURRENT APPLICATION NUMBER: US/09/103,840A
NUMBER OF SEQ ID NOS: 2
NUMBER OF SEQ ID NOS: 2
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 4411529
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
OTHER INFORMATION: H37Rv
US-09-103-840A-1

Qy 676 IaleuAspGlySerAlaValAlaGlyAla 685
Db 1809396 CGATCGCTGCCCGGACGACGCTGGCGCT 1809424

RESULT 13
US-09-252-991A-811
Sequence 811, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
PRIOR FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 811
LENGTH: 1506
TYPE: DNA
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-811

Alignment Scores:
Pred. No.: 5,34e-48 Length: 1506
Score: 563.00 Matches: 167
Percent Similarity: 46.56% Conservative: 77
Best Local Similarity: 31.87% Mismatches: 214
Query Match: 15.76% Indels: 66
Gaps: 15

US-10-089-514-2 (1-686) x US-09-252-991A-811 (1-1506)

Qy 172 IleGlySerAspPheGlyArgGluIleMetAlaSerPheArgAspLeuAlaHis 191
Db 52 CTGCCCCCGATGGCTACACACCGCATCCCGCTGCTTCGAGACCTTGC----- 102

Qy 192 HisArgAlaArgArgHisGlyAlaAspSerProTyrGluLeuHisValArgArgValAsp 211
Db 103 -----GACTTCGACACGCGCGCTGCGATCTACCTGAAGCTGGCGGAC 144

Qy 212 ValLeuProAspAlaGluGluValArgArgGlyCysLeuProGlyGluGlyThrPhe 231
Db 145 GCG-----CGGAACCTCTAC 159

Qy 232 TrpLeuAspSerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAsp 251
Db 160 CTCTGGAGTCTCGTCAGGCGGCGAGAAATGCGGCGCTATTGTCATCATCGGC----- 213

Qy 252 ArgGlyProLeuAlaGluTyrLeuThrTyrArgValAlaAspGlyValValSerValArg 271
Db 214 -----CTGCCGTGCGACGCGTGTGCGGAGTCTACGACCATCAAGTCGGATC--- 261

Qy 272 GlySerAspGlyThrThrThrArgThrArgArgPhePheAsn----- 286
Db 262 AGCATCATGGCATGGAACCGAGCGC-----TTTCATTCGCGCGACCGCGCTG 309

Qy 287 ---TyrLeuGluGluGluGluGluArgArgArgValProValAlaProGluLeuProPhe 305
Db 310 GCTTTCCTCGAGGCTTCAAGCGCGCTACCAAGGTGCCACCGTCCGCGCTTCCCA--- 366

Qy 306 GluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyr-----GluLeu 319
Db 367 CGTTTCATGGCGCGCTGTGCTGCTACTTCGTTTACATGCGTGGCTACGTTGAAAA 426

Qy 320 LysAlaGluThrThrGlyAspProAlaHisArgSerPro-----HisProAspAlaAla 337
Db 427 CGCCTGGCACCTCTCGAACCG-----GACCCGCTGGGCAACCGGATATCTCTG 477

Qy 338 PheLeuPheAlaAspArgAlaIleAlaLeuAspHisGluGluCysCysTyrLeuLeu 357
Db 478 TTGATGTCGTCGATGCGGATGCGTATTCGACACCTGGCCGGGAAGATCCACGCCATC 537

Qy 358 AlaLeu-----AspArgArgGlyHisAspAspGlyValAlaArgAlaTrpLeuArg 373
Db 538 GTCTCGCGCATCCCTCCGAGAAATGCTTACGAGCGCGC---GAGCAGCTCTGAG 594

Qy 374 GluThrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrProAla 393
Db 595 GACCTGCTGAGCGCTG-----CGCCAGCGCATGCCCGCGCTCGCGC 639

Qy 394 MetValPheGlyLeuProGluAlaAlaGlyPheGlyProLeuAlaArgAlaArgHis 413
Db 640 CTGACCTC-----GAGCGCGCCAGGCTGTGAGCGCGCTTCTGCTGCACTTC 690

Qy 414 AspLysAspAlaTyrLeuLysArgGlyLeuAspGlyCysLeuLysGluLeuArgAsnGlyGlu 433
Db 691 ACCCGGAGCATATGAAACCGGTAGGAAGATCAAGGATCATCTTGGCGCGGAC 750

Qy 434 SerTyrGluLeuLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuPro 453
Db 751 TGCATGAGGTGTCGCTGCGCGCATGCTCATGCAATTCAGGCGCGCGCCATCGAC 810

Qy 454 LeuTyrSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPhePro 473
Db 811 CTGTACCGCGCGCTGCGCTGTTTCAATCCGACCGCTTACATGTTCTTCAACTTCGCGC 870

Qy 474 GluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGly 493
Db 871 GACTTCATTCCTGTCGGGACGCTGCGGAGCTGTCTGACCGGTC---GAGGATGGCTG 927

Qy 494 ValGluSerLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGlu 513
Db 928 GTGACCGTGGCGCGATGCGCGCTACCGGCGCGCGGATCAACAAGCGCGGACCTG 987

Qy 514 ArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArgAlaGluLeuMetIleVal 533
Db 988 GCACTGAGCAGGATGCTGCTGTCGACCGCCAAAGAAATGCGCAGACCTGATGTGATC 1047

Qy 534 AspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArg 553
Db 1048 GACTTGGGCGCAACACGCGGCGGCTGTCCGATATCGCGCGGTGAAGTCAACCGAA 1107

Qy 554 LeuPheGluValGluThrTyrAlaProValHisGluLeuValSerThrIleArgGlyArg 573
Db 1108 AAATGCGTATGCAAGCTTATCTCAACGTCAATGCAATGTCATGCTGCAACGTCACCGGCA 1167

Qy 574 LeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyLysMet 593
Db 1168 TTCGCGGAGGCTCAGCGCGATGAGCGCGCTGCGGCGCATTTCTGCGGCGGCACTCTA 1227

Qy 594 ThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeuGluGluProArg 613
Db 1228 TCCGCGCGCCCAAGTCCGCGCATGAGATATGACGAGCTGAGCCGCTCAAGCT 1287

Qy 614 GlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle 633
Db 1288 GAGTCTACGCGCGCGCGCTGCTACCTGAGTGAAGGACGGAACATGACACCGCAT 1347

Qy 634 ValIleArgThrIleValLeuAlaAspGlyGluAlaGluPheGlyValGlyAlaIle 653
Db 1348 GCCATCCGACCGCGGTGATCAAGACGCTGAATCTCAGCTGAGCGCGCGCGCTATC 1407

Qy 654 ValSerLeuSerAspGluGluGluGluPheThrGluThrValValLysAlaArgAlaMet 673
Db 1408 GTTGCAGACTCGTGTCCCGCGCTGAGTGGAGAAACATCAACAAGCGCGGAGATG 1467

Qy 674 ValThrAlaLeu 677
Db 1468 TTCGCGCGCGT 1479

RESULT 14

```
US-09-328-352-1525
; Sequence 1525, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328.352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 1525
; LENGTH: 1503
; TYPE: DNA
; ORGANISM: Acinetobacter baumannii
US-09-328-352-1525

Alignment Scores:
Pred. No.: 1,75e-47 Length: 1503
Score: 558.00 Matches: 162
Percent Similarity: 47.38% Conservative: 91
Best Local Similarity: 30.34% Mismatches: 203
Query Match: 15.62% Indels: 78
DB: Gaps: 15

US-10-089-514-2 (1-686) x US-09-328-352-1525 (1-1503)

QY 181 MetAlaSnPheArGpLeuAlaLeuAlaHis-----HisArgAla 194
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 19 TTAGCGCAATTTCCAGCACTTTAAGCGTCTGCTGTTACATCTATTCTGTTATTCGCA 78

QY 195 ArgArGhisGlyAlaAspSerProTyGluLeuHisValArgArgValaAspValLeuPro 214
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 79 CGTTACCGGATACAGAAACCCATTCGTGTTTCTCGTTTAAGAT-----129

QY 215 AspaIaGluGluValaArgArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAsp 234
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 130 -----CAAAACAGCGCTCTTATTGAG 153

QY 235 SerSerSerValLeuGluGlyAlaSerArgPheSerPheLeuAspArgGlyPro 254
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 154 TCTGTTAAGGGGTGAAACTGGCGCGCTACTCTATGATGATG-----198

QY 255 LeuAlaGluTrpLeuThrTrpArgValaAlaAspGlyValaSerValaArgGlySerAsp 274
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 199 TTAGCGCAATCCACAGTTTTCCTGTAACGCTGCTTTATCTATACAAACATGGTGAC 258

QY 275 GlyThrThrThrArgThrArg-----ArgProPhePheAsnTyTrpLeuGluGluLeu 292
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 259 GGATCACTAAACACGCAAAATTTGCTAGACCA--TTCCAAATATATCCGTGAATCCAA 315

QY 293 GluArgArgArgValaProValAla-----ProGluLeuProPheGluPheAsnLeu 309
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 AAACAATTTTAAAGTCTCTACGGCTAAACTCTTACCAACATTACCA--ACCTTACGGCG 372

QY 310 GlyTrpValGlyTrpLeuGlyTyTrpGlu-----LeuVala 321
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 373 GCGTGTGGGTATTTGGGCTACGATGCTGTCCGCTATATCGAGCCACCTTTAAAGAT 432

QY 322 GluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAla 341
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 433 GTACCTACGGCTGATCCGATT-----ACGCTACCAAGATTATGTTGATGCTCTCA 483

QY 342 AspArgAlaAlaLeuAlaSerHisGlnGluGlyCysGlySerTyTrpLeuAlaLeuAspArg 361
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 484 AAGACACTCATGTTGTTTGAACAATCTTAAAGATACGCTATTTTAAATTG-----534

QY 362 ArgGlyHisAspArgGlyAlaArgAlaTrpLeuArgGluThrAla-----376
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 535 -----CATCGGATTAAGACACAGATATGCTTATATAAGACGCTACGCAAAATTAAGAT 588

QY 377 -----GluThrLeuThrGlyLeuAlaValaArgAlaProAlaGluPro---ThrProAla 393
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 589 CAATTAGAACAGTTGTTGGCGACTCCAGTTAGTTGCAAGCCCAACCATATCGCT---645
QY 394 MetValPheGlyIleProGluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHis 413
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 -----CCGCATTTTGAATCAATTAAT 666

QY 414 AspLysAspAlaTyTrpLeuLysArgIleAspGluCysLeuLeuGluIleArgHisGlyGlu 433
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 GGTAAGCAAAATTTCTTAAAGACAGTAGAAGAGTTTAAAGATATATCCGTACAGCGCAT 726

QY 434 SerTyGluIleCysLeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuPro 453
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 GTGATCGAGTTGTATCTGCGACGCGTATGTTGCTTATGTTGCTATTCAGAGAGCTTTACG 786

QY 454 LeuTySerAlaLeuArgAlaIleSerProValProTyGlyAlaLeuLeuGluPhePro 473
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 GTTTACCGTCATTAAGTATTAATTCATCACCTTATCTATCTGTTGTTCAAGCGCA 846

QY 474 GluLeuSer-----ValLeuSerAlaSerProGluArgPheLeuThr 487
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 847 ACGATTACTGATTAATAAACCATTTTCATATTTGCTCATCACCGGAA--ATTTTATCT 903

QY 488 IleGlyAlaAspGlyGlyValaGluSerLysProIleLeuGlyThrArgProArgGlyGly 507
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 904 CGTTTAAAGAACGGTATTTGCTACAGTTTCAGCTTGGCAGGAAGTACGACCGCGTAAA 963

QY 508 ThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArgAla 527
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 ACTTAAAGAAAGATATATACATTTAGAAAAGATTTCTGTCGTGATGAAAAGAGATGCT 1023

QY 528 GluAsnLeuMetIleValaAspLeuValaArgAsnAspLeuAsnSerValCysAlaIleGly 547
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 GAACATTATATGCTGATGATGATCTTGGCGAAACGATGTAGCGCGCTATCGAAAATAGCT 1083

QY 548 SerValHisValProArgLeuPheGluValaGluThrTrpAlaProValHisGlnLeuVal 567
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 AAGATCCAAATCCAGATCGGATCAAAATGTGATGTAGCGCTATTCACATGTATGATTTGT 1143

QY 568 SerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValaArgAlaAla 587
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1144 TCAAATGTACAAAGTGAAGTGAAGTGTGATGATATCGATGCACTTGATATTTAAAGCCACC 1203

QY 588 PheProGlyGlySerMetThrGlyAlaProLysLysArgThrMetGluIleIleAspArg 607
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1204 TTTCCGGCAGGAAGCTTATCAGGTGCCCCCAAAATTCGTGCAATGGAATATTGATGAA 1263

QY 608 LeuGluGluGlyProArgGlyValaTySerGlyAlaLeuGlyTrpPheAlaLeuSerGly 627
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1264 GTAGAACCTGTGAAGAGGGAGATTGTTGGCGGGCGTGTGTTATTTGGGATGGCATGCT 1323

QY 628 AlaAlaAspLeuSerIleValIleArgThrIleValaLeuAlaAspGlyGluAlaGluPhe 647
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1324 GAAATGGATATATGTCATATTCATTCCTGTTGTTATTCGTCGCAAAAAGGTATGTGA 1383

QY 648 GlyValaGlyAlaIleValaSerLeuSerAspArgGlnGluGluPheThrGlyThrVal 667
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1384 CAGCGTGTGACAGGGCTAGTGTGTGACTCAAAATCCAGATTCGATGCGAATGAAACTCA 1443

QY 668 ValLysAlaArgAlaMetValThrAlaLeuAspGlySerAla 681
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1444 ATTAAAGCTCGCGCAGTGATCAAAAGCGTGTATATATCATCA 1485

RESULT 15
US-08-403-852D-9
; Sequence 9, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanché, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
```


This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 1509.69 Seconds

(without alignments)
3137.453 Million cell updates/sec

Title: US-10-089-514-2

Perfect score: 3573

Sequence: 1 MRTLIDNYSFTHTLFOYI.....VKARAWVTALDGSAAVAGAR 686

Scoring table:

BLOSUM62

Xgapop 10.0, Xgapext 0.5

Ygapop 10.0, Ygapext 0.5

Fgapop 6.0, Fgapext 7.0

Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame+g2n.model -DEV=rlh

-Q=cg2_1/USPTO.epool_n/US10089514/runat_04102005.105745.8055/app.query.fasta_1.1621

-DB=Published.Applications.NA -QFMT=fastap -SUFFIX=trmb -MINMATCH=0.1

-LOOPEXT=0 -LOOPEXT=0 -UNITS=bases -START=1 -END=1 -MATRIX=blosum62

-TRANS-human40.csl -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100

-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0

-MAXLEN=200000000 -USER=US10089514 @CGN_1_1_920 @runat_04102005.105745.8055

-NCPU=6 -ICPU=3 -NO_MMAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100

-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5

-Fgapop=6 -Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

Database:

Published Applications NA:*

1: /cg2_6/prodata/1/pubna/US07_PUBCOMB.seq:*

2: /cg2_6/prodata/1/pubna/PCT_NEW_PUB.seq:*

3: /cg2_6/prodata/1/pubna/US06_NEW_PUB.seq:*

4: /cg2_6/prodata/1/pubna/US06_PUBCOMB.seq:*

5: /cg2_6/prodata/1/pubna/PCTUS_NEW_PUB.seq:*

6: /cg2_6/prodata/1/pubna/PCTUS_PUBCOMB.seq:*

7: /cg2_6/prodata/1/pubna/US08_NEW_PUB.seq:*

8: /cg2_6/prodata/1/pubna/US08_PUBCOMB.seq:*

9: /cg2_6/prodata/1/pubna/US09_PUBCOMB.seq:*

10: /cg2_6/prodata/1/pubna/US09_PUBCOMB.seq:*

11: /cg2_6/prodata/1/pubna/US09C_PUBCOMB.seq:*

12: /cg2_6/prodata/1/pubna/US09_NEW_PUB.seq:*

13: /cg2_6/prodata/1/pubna/US10_PUBCOMB.seq:*

14: /cg2_6/prodata/1/pubna/US10_PUBCOMB.seq:*

15: /cg2_6/prodata/1/pubna/US10C_PUBCOMB.seq:*

16: /cg2_6/prodata/1/pubna/US10_PUBCOMB.seq:*

17: /cg2_6/prodata/1/pubna/US10C_PUBCOMB.seq:*

18: /cg2_6/prodata/1/pubna/US10_PUBCOMB.seq:*

19: /cg2_6/prodata/1/pubna/US10C_PUBCOMB.seq:*

20: /cg2_6/prodata/1/pubna/US10_PUBCOMB.seq:*

21: /cg2_6/prodata/1/pubna/US10C_PUBCOMB.seq:*

22: /cg2_6/prodata/1/pubna/US11_NEW_PUB.seq:*

23: /cg2_6/prodata/1/pubna/US11_PUBCOMB.seq:*

24: /cg2_6/prodata/1/pubna/US11_NEW_PUB.seq:*

25: /cg2_6/prodata/1/pubna/US60_NEW_PUB.seq:*

26: /cg2_6/prodata/1/pubna/US60_PUBCOMB.seq:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3573	100.0	2061	20	US-10-472-587-1
2	3127.5	87.5	3305	14	US-10-126-927-68
3	3127.5	87.5	12391	14	US-10-126-927-68
4	1766	44.7	438203	21	US-10-819-386A-1
5	1596.5	44.7	2196	15	US-10-156-761-1167
6	1596.5	44.7	9025608	15	US-10-156-761-11
7	1487	41.6	1860	9	US-09-738-626-1111
8	1487	41.6	2005	21	US-10-494-675-25
9	1487	41.6	3309400	9	US-09-738-626-11
10	1350	37.8	2046	22	US-10-501-282-17
11	1350	37.8	1754382	22	US-10-501-282-17
12	1210.5	33.9	3023	19	US-10-437-963-3537
13	1068.5	29.9	2523	17	US-10-282-122A-17825
14	1001	28.0	1906	17	US-10-282-122A-16977
15	834	23.3	684707	17	US-10-398-221-9
16	834	23.3	3011208	17	US-10-398-221-9
17	697.5	19.5	1066	18	US-10-424-599-33601
18	623	17.4	1362	17	US-10-282-122A-30135
19	585	16.4	9025608	15	US-10-156-761-1
20	579.5	16.2	1377	17	US-10-282-122A-41081
21	568	15.9	1377	17	US-10-282-122A-41824
22	566.5	15.9	1338	17	US-10-282-122A-32104
23	566.5	15.9	1479	15	US-10-156-761-6148
24	563	15.8	1323	17	US-10-282-122A-19610
25	562.5	15.7	661	20	US-10-425-115-81629
26	561.5	15.7	1365	17	US-10-282-122A-38883
27	561	15.7	1440	17	US-10-282-122A-38883
28	557.5	15.6	1365	17	US-10-282-122A-39512
29	556	15.6	645	20	US-10-716-803-9
30	544.5	15.2	1257	17	US-10-282-122A-33856
31	539.5	15.1	1333	17	US-10-282-122A-23903
32	538.5	15.1	5721	17	US-10-398-221-3850
33	534.5	14.9	1337	17	US-10-282-122A-76794
34	531.5	14.9	495269	17	US-10-398-221-8
35	531.5	14.9	3011208	17	US-10-398-221-8058
36	526.5	14.7	1362	17	US-10-282-122A-6689
37	519.5	14.5	2073	22	US-10-450-763-29714
38	509	14.2	1425	17	US-10-282-122A-32780
39	504	14.1	1284	9	US-09-974-300-482
40	500	14.0	5396	17	US-10-398-221-3849
41	499.5	14.0	1898	22	US-10-450-763-1954
42	498.5	14.0	2256646	19	US-10-470-565-1
43	498	13.9	1518	19	US-10-375-039-15
44	497.5	13.9	2731748	19	US-10-297-465A-1
45	489	13.7	1794	18	US-10-275-026A-173

ALIGNMENTS

RESULT 1
US-10-472-587-1
; Sequence 1, Application US/10472587
; Publication No. US20040214274A1
GENERAL INFORMATION:
APPLICANT: YANAI, Koji
APPLICANT: SUMIDA, Naomi
APPLICANT: WATANABE, Manabu
APPLICANT: MORIYA, Tatsuki
APPLICANT: MURAKAMI, Takehi
TITLE OF INVENTION: Transferrin Producing Substrate P1022 Derivatives, Methods for
FILE REFERENCE: 2003-1302A/MMC/00144
CURRENT APPLICATION NUMBER: US/10/472,587
CURRENT FILING DATE: 2003-09-22
PRIOR APPLICATION NUMBER: 82227/2001

Pred. No. is the number of results predicted by chance to have a

```
/ PRIOR FILING DATE: 2001-03-22
/ NUMBER OF SEQ ID NOS: 40
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1
/ LENGTH: 2061
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(2058)
/ US-10-472-587-1

Alignment Scores:
Pred. No.: 0 Length: 2061
Score: 3573.00 Matches: 686
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0

US-10-089-514-2 (1-686) x US-10-472-587-1 (1-2061)

QY 1 MetArgThrIleuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 1 ATGCGCAGCGCTTCGATCGACAACACTACGACTCGTTCACTCCAACTGTTCCAGTACATC 60

QY 21 G1G1uLaIaThrG1yGlnPProProValValaProAsnAspAlaAspTTPSerArgLeu 40
Db 61 GGGAGGCGCACCGGGCAACCCCGCTGCTGTCGCCAACGACGCCGCTGTCGGGCTG 120

QY 41 ProValG1uAspPheAspAlaIleValIleSerProG1ySerProAspArgG1u 60
Db 121 CCGCTCGAGAGACTTCGACGCGATCGTGTCTCCCGGGCCCCCGGACCGCGAGAA 180

QY 61 ArgAspPheG1yIleSerArgArgAlaIleThrAspSerG1yLeuProValIleuG1yVal 80
Db 181 CGGAGACTTCGGAAATCAAGCCCGGGCGATCACCAGACGAGCGGCTCGCTCGGGGT 240

QY 81 CysLeuG1yHisGlnG1yIleAlaGlnLeuPheG1yG1yThrValG1yLeuAlaProG1u 100
Db 241 TGGCTCGGCCACCAAGGCGATCGCCAGCTTCGGCGGAACCGTGGCTCGGCCGGA 300

QY 101 ProMetHisG1yArgValSerG1uValArgHisThrG1yG1uAspValPheArgG1yLeu 120
Db 301 CCCATGACGCGCGGCTCCGAGGTCCGSCACACGCGCGAGACGCTTCGCGGGCTC 360

QY 121 ProSerProPheThrAlaValArgTyrHisSerIleuAlaIleThrAspLeuProAspG1u 140
Db 361 CCTCGCGCTTCAACCGCGCTGCGCTACCACTCCCTGGCGCCACCGACTCCCGGAGAG 420

QY 141 LeuG1uProIleuAlaTTPSerAspAspG1yValIleMetG1yLeuArgHisArgG1uLys 160
Db 421 CTGGAAACCTTCGCTGAGCGAGCAGACGAGGTGCTATGGGCTGCGCGCACCGGAGAG 480

QY 161 ProIleuTyrG1yValGlnPheHisProG1uSerIleG1ySerAspPheG1yArgG1uIle 180
Db 481 CCGCTGGGGGCGTCCAGTTCACACCGGAGTCCATCGGACGACTTCGCGCGGAGATC 540

QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisHisArgAlaArgArgHisG1yAlaAsp 200
Db 541 ATGGCCAACTTCGCGGACCTCGGCTTCGCCACACCGGGCACCGGGCGCACGGGGCGAC 600

QY 201 SerProTyrG1yLeuHisValaArgArgValaAspValIleuProAspAlaG1uG1uValArg 220
Db 601 TCCCGCAACCACTCCACGTCGCGCGCTGACAGTGTGCGGACCGCGAAGGATACGC 660

QY 221 ArgG1yCysIleuProG1yG1uG1yThrThrPheTyrLeuAspSerSerValIleuG1u 240
Db 661 CGGCGCTGCTCGCGCGGAGGACCACTTCGCTGGAGACAGACTTCGCTCTTGAA 720

QY 241 G1yAlaSerArgPheSerPheLeuG1yAspAspArgG1yProIleuAlaG1yTyrLeuThr 260
Db 721 GGGGCTCGGCTTCCTCTCTCTCTCGGAGACGCGGGCGGCTCGCGGAGTACCTCACC 780
```

```
QY 261 TyrArgValAlaAspG1yValValSerValaArgG1ySerAspG1yThrThrThrArgThr 280
Db 781 TACCGGTGCGGACGAGCGTCTTCCTCCGCGGCTCCGACGACCAACGACCGGAGC 840

QY 281 ArgArgProPhePheAsnTyrIleuG1uGlnIleuG1uArgArgValaProValAla 300
Db 841 CGGCGCGCTTCCTTCAACTACCTGGAGGAGCAGCTCGAAACGCGGACGGGTCCCGTCCGC 900

QY 301 ProG1uLeuProPheG1uPheAsnLeuG1yTyrValG1yTyrLeuG1yTyrTyrIleuLys 320
Db 901 CCGGAATCGCCCTTCGAGTTCAACTCGGCTACGTCGGCTACCTCGGCTACGAGTGAAG 960

QY 321 AlaG1uThrThrG1yAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db 961 GCGGAGAACCCGCGGACCCCGGACCGGTTCGCGACCCGACCGCGCTTCCTTC 1020

QY 341 AlaAspArgAlaIleAlaLeuAspHisGlnG1uG1yCysArgTyrIleuLeuAlaLeuAsp 360
Db 1021 GCGGACGCGGCACTGCGCTCGAACCAAGAGGCTGCTACCTGCTGCGCCCTCGAC 1080

QY 361 ArgArgG1yHisAspAspG1yAlaArgAlaTTPLeuArgG1uThrAlaG1uThrThr 380
Db 1081 CGCGGGGCGCACGACGACGCGCGCTGCTGCGGAGACGCGCGGACCTTCACC 1140

QY 381 G1yLeuAlaValArgAlaProAlaG1uProThrProAlaMetValPheG1yIleProG1u 400
Db 1141 GGCCTGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1200

QY 401 AlaAlaAlaG1yPheG1yProIleuAlaArgAlaArgHisAspLysAspAlaTyrIleuLys 420
Db 1201 GGGGGGCGGCTTGCGGCCCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260

QY 421 ArgIleAspG1uCysLeuLysG1uIleArgAsnG1yG1uSerTyrG1uIleCysLeuThr 440
Db 1261 CGCATCGAGAGTGTCTTCAGAGAGATCCGCCAAGCGGAGTCTTACGAGATTCGCTGACC 1320

QY 441 AsnMetValThrAlaProThrG1uAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
Db 1321 AACATGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1380

QY 461 IleSerProValProTyrG1yAlaLeuLeuG1uPheProG1uLeuSerValIleuSerAla 480
Db 1381 ATCAGCCCGCTCCGTAAGCGCGCTCGTCAAGTTCGCCGAATCGGTGCTGAGCGCG 1440

QY 481 SerProG1uArgPheLeuThrIleG1yAlaAspG1yG1yValG1uSerLysProIleLys 500
Db 1441 TCGCCCGAGCGGTTCTTCACGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500

QY 501 G1yThrArgProArgG1yG1yThrAlaG1uG1uAspG1uArgLeuArgAlaAspLeuAla 520
Db 1501 GGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560

QY 521 G1yArgG1uLysAspArgAlaG1uAsnLeuMetIleValaAspLeuValaArgAsnAspLeu 540
Db 1561 GCGCGGGAAGAAGACCGGCGCGGAACTGATGATGTGACCTGTGCGGACCAACACTC 1620

QY 541 AsnSerValCysAlaIleG1ySerValHisValProArgLeuPheG1uAlaG1uThrTyr 560
Db 1621 AACAGCGGTGCGCGCATCGGCTCCGTCACGTCGCGCGCGCTCTTCGAGTGAAGACTC 1680

QY 561 AlaProValHisGlnLeuValSerThrIleArgG1yArgLeuArgPProG1yThrSerThr 580
Db 1681 GCGCGCTGACCAAGTGTGTGACCACTCGGGGAGCGGCTCGCGCGCGCGCGCGCGCG 1740

QY 581 AlaAlaCysValArgAlaAlaPheProG1yG1ySerMetThrG1yAlaProLysLysArg 600
Db 1741 GCGGCTGCTGACGCGCGCGCTTCGCGCGGCTCCTCAATACCGGCGCGCGCGCGCGCG 1800

QY 601 ThrMetG1uIleIleAspArgLeuG1uG1yProArgG1yValTyrSerG1yAlaLeu 620
Db 1801 ACCATGAGATCATGACCGCTCGAGGAGGACCGCGCGGCGGCTCACTCCGGGCGCTC 1860
```

QY 621 G1YrPheh1aleuSerG1yAla1aAsp1euSer1leVal1leArgThr1leVal1leu 640
DB 1861 GGATGTTCCGCTTCAAGCGGCGCCGACCTTCAAGATCGTCATCCGACCATGCTGCTG 1920
QY 641 AlaAspG1yGln1aGluPheG1yVal1G1yAla1leVal1Ser1euSerAspG1yGlu 660
DB 1921 GCCGACGGCCAGGGGAGTTTCGGGTGCGGGGCGAGTCGTCTCCCTCCGACCAAGAG 1980
QY 661 G1G1uPheThrG1uThrVal1Val1yAlaArg1aMetVal1Thr1a1euAspG1ySer 680
DB 1981 GAGGAGTTTCAACGAGACCGGTGTAAAGCCCGCCCATGTCACCGCTCGACGGGAGC 2040
QY 681 AlaVal1aG1yAlaArg 686
DB 2041 GCCGTGGCGGGCGCCCGA 2058
RESULT 2
US-10-126-927-68
; Sequence 68, Application US/10126927
; Publication No. US20030082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Meh1, Ryan A
; APPLICANT: Pastinak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
; LENGTH: 3305
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-126-927-68
Alignment Scores:
Pred. No.: 1,3e-301 Length: 3305
Score: 3127.50 Matches: 618
Percent Similarity: 91.40% Conservative: 9
Best Local Similarity: 90.09% Mismatches: 42
Query Match: 87.53% Indels: 17
DB: 14 Gaps: 3
US-10-089-514-2 (1-686) x US-10-126-927-68 (1-3305)
QY 1 MetArgThr1eu1leu1leuAsp1eu1TyAsp1eu1Ser1Phe1Thr1Ser1eu1Phe1Gln1Ty1le 20
DB 1 ATGGGCGCGCTTCAAGTCACTACGACTCGTTCCAGCAACCTGTTCCAGTACATC 60
QY 21 G1yG1u1a1aThrG1yGln1Pro1a1Val1a1Pro1a1aAsp1a1aSer1Phe1Arg1eu 40
DB 61 GGGAGGCGCACCGGGGAGCCCGCC---GTGTCGCCAAGCGCGCATGCTGCGCGCTG 117
QY 41 ProVal1G1uAsp1PheAsp1a1leVal1a1Ser1ProG1ySer1ProAspArg1Glu 60
DB 118 CCCCTCAGAGACTTCGACGGGATGTCGTGCTCCCGGGCCCGGCAAGCCCGGAGAA 177
QY 61 ArgAspPheG1y1leSer1Arg1a1le1eThrAspSerG1y1euProVal1leuG1yVal 80

DB 178 CGGAGCTTCGAGATCACCCCGGCGGATACCGACAGCGCGCTGCCCTTCGCGCTC 237
QY 81 Cys1euG1yH1aG1nG1y1le1a1G1n1euPheG1yG1yThrVal1G1y1euAlaProG1u 100
DB 238 TGCTCGGCCACCAAGGAGCATGCGCAG-----CTTCGCGGAGAA 276
QY 101 ProMetH1aG1yArgVal1SerG1uVal1ArgH1eThrG1yG1uAspVal1PheArgG1y1eu 120
DB 277 CCCATGACGGCGGGGCTCCAGAGTCCGGCACACCGGAGAGGAGCGTCTCCGGGGCTC 336
QY 121 ProSerProPheThrAlaVal1Arg1yH1eSer1euAla1aThrAsp1euProAspG1u 140
DB 337 CCTCGCGCTTCAACCGCGCTGCGCTACCTCCCTGCGCCGACCACTCTCCCGACGAG 396
QY 141 LeuG1uPhe1u1a1aThrPhe1a1aAspArg1yVal1MetG1y1euArgH1aArgG1u1y 160
DB 397 CTGAAACCTTCGCTCGAGCGACGAGCGGCTGTCAAGGACCTGCGGACCGGAGAG 456
QY 161 ProLeuThrP1yVal1GlnPheH1aProG1uSer1leG1ySerAspPheG1yArgG1u1le 180
DB 457 CCGCTGATGGGCGCTTCCAGTTCACCGGAGTCCATCGGACGACATTCGCGCGGAGATC 516
QY 181 MetAlaAspPheArgAsp1euAla1euAlaH1eH1aArg1a1aArgArgH1aG1yAlaAsp 200
DB 517 ATGGCCAACTTCGCGGACCTCGCCCTGCGCCACACCGGCGACGTCGCGACCGCGGAC 576
QY 201 SerProTyG1u1euH1a1aVal1a1ArgVal1a1AspVal1euProAsp1a1aG1uVal1a1Arg 220
DB 577 TGGGGCTACCAACTCCACGCGCGCGCTGCGAGCTGTCGCGGACCGCGGAAGAGTACGC 636
QY 221 ArgG1yCys1euProG1yG1uG1yThrThrPheThr1euAspSer1SerVal1leuG1u 240
DB 637 GCGGCTGCTGCCCGCGCGGCGGCGGCGGACCTTCGCTGAGACAGCACTTCGCTTGA 696
QY 241 G1yAlaSerArgPheSerPhe1euG1yAspAspArgG1yPro1euAlaG1uTy1euThr 260
DB 697 GCGGCTCGCGCTTCTCTCTCTCGGAGAGACCGGCGCGCTCGCGGAGTACCTCAC 756
QY 261 TyrArgVal1a1aAspG1yVal1a1SerVal1a1ArgG1ySerAspG1yThrThrArgThr 280
DB 757 TACCGGCTCGCGACGCGCGCTGCTCGGCGGCTCGGCGGCTCGGCGGACGACGACGCGGAG 816
QY 281 ArgArgProPhePhe1euThr1euG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uG1uG1u 300
DB 817 GCGGCGACCTTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACT 876
QY 301 ProG1u1euProPheG1uPhe1eu1euG1yTy1a1G1yTy1euG1yTy1euG1y 320
DB 877 CCGGACCTGCGCTTCAAGTTCAACTCGGCTACGTCGCTACCTCGGCTACGAGTGAAG 936
QY 321 AlaG1uThrThrG1yAspProAlaH1a1aArgSerProH1a1a1aPhe1euPhe 340
DB 937 GCGGAGACCAACCGGCGACCGCGAGTACCGGCGGACCGGCGGCTCTCTTC 996
QY 341 AlaAspArgAla1le1a1euAspH1a1aG1nG1yCysG1yTy1eu1euAla1euAsp 360
DB 997 GCGGACCGCGCATTCGCGCTCGACCAACGAGAAAGGCTGCTACCTGCTGCGCTCGAC 1056
QY 361 ArgArgG1yH1a1aAspAspG1yAlaArgAla1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 380
DB 1057 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1116
QY 381 G1y1euAlaVal1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 400
DB 1117 GCGCTGGCGGTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1176
QY 401 AlaAla1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 420
DB 1177 GCGGCGGCGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1227
QY 421 Arg1laAspG1u1y1eu1y1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a1a 440
DB 1228 -----TCGCGCTTCGCGGAGGAGTGTGTAAGATCTGCTGAC 1269

```
QY 441 AsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAla 460
DB 1270 AACATGGTCAACCGCGGACCGAGCGAGCGGCGCTGCGCTACTCCGCGTGGCGGC 1329
QY 461 IleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValIleSerAla 480
DB 1330 ATCAGCCCGCTCCGCTGCGGCGCTGCTGAGTTCCTCCGAGTGTCCGAGTCTCAGCGCC 1389
QY 481 SerProGluPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysProIleLys 500
DB 1390 TCCCGGAGGCGTTCCTCAGATCGGCGCGGACGCGCGCTCCAGTCCAGGCCATCAG 1449
QY 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
DB 1450 GGGACCGCGCCCGGGGCGGACCGGCGAGAGAGACAGAGGCGCTCCGCGCGGACCTGGCC 1509
QY 521 GlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
DB 1510 GCGCGGAGAGAGACCGGCGCGGAGAACCTGATGATGTCGACCTGTCCGCAACGACCTC 1569
QY 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
DB 1570 AACAGGCTGCGCGGATCGGCTCCGTCACGTCGCGCTTCCTCAGAGTGGAGACCTC 1629
QY 561 AlaProValHisGluLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThr 580
DB 1630 GCGCCCGTGACACGCGTGTGACACCATCGGGGAGACGCGTCCGCGCGGACCAACAC 1689
QY 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArg 600
DB 1690 GCGCGCTGCTAGCGCGGCTTCCTCCGCGGCTCCATGACCGCGCGGACCAAGAGCGA 1749
QY 601 ThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGlyAlaLeu 620
DB 1750 CCATGGAGATCATTCACCGCTCGAGAGAGAGGCCCGGGGCGCTTACCGGGGCGCTC 1809
QY 621 GlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
DB 1810 GATGTGTTCCCTCAGCGCGCGCGGACCTCAGCATCGTATCCGACCATCGTGTCTG 1869
QY 641 AlaAspGlyGluAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGluGlu 660
DB 1870 GCGGACCGCGCGGCGGATTCGCGTGGCGGCGGCGATCGTTCCTCCGACCGAGAG 1929
QY 661 GluGluPheThrGluThrValIleLysAlaArgAlaMetValThrAlaLeuAspGlySer 680
DB 1930 GAGGAGTTCAGGACGCGGTGTCAGAGCCCGCGCATGTGTCACTCCCTTCGACGCGC 1989
QY 681 AlaValAlaGlyAlaArg 686
DB 1990 GCAGTGGCGGCGCGCGCA 2007

RESULT 3
US-10-126-927-67
; Sequence 67, Application US/10126927
; Publication No. US20030082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pasternak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
```

```
; CURRENT FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 12391
; TYPE: DNA
; ORGANISM: Plasmid pSC101, Streptomyces venezuelae papABC
US-10-126-927-67

Alignment Scores:
Pred. No.: 7,13e-301 Length: 12391
Score: 3127.50 Matches: 618
Percent Similarity: 91.40% Conservative: 9
Best Local Similarity: 90.09% Mismatches: 42
Query Match: 87.53% Indels: 17
DB: 14 Gaps: 3

US-10-089-514-2 (1-686) x US-10-126-927-67 (1-12391)
QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGluTyrIle 20
DB 24 ATGCGACGCTTCTGATCAACAATAAGACTGTTCAACCAAGACCTGTTCAATACATC 83
QY 21 GlyAlaAlaThrGlyGluProProValValIleProAsnAspAlaAspTyrSerArgLeu 40
DB 84 GCGAGAGCACCGGCGAGCGCCCC---GTCTGTCCCAAGACCGGACATGTGCGGCTG 140
QY 41 ProValGluAspPheAspAlaIleValIleSerProGlyProGlySerProAspArgGlu 60
DB 141 CCCTCGAGAGACTTCAGACCGATCGTGTCTCCCGGCGCGGACCGGACCGGAGAA 200
QY 61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
DB 201 CCGGACTTGAGATCAAGCGCGCGGATCACCGAGCGGCTGCTCCCTCGGCGCTC 260
QY 81 CysLeuGlyHisGluGlyIleAlaGluLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
DB 261 TGCTCGGCGCACGAGGCGATTCGCCAG-----CTTCGCGCGAA 299
QY 101 PromethiGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
DB 300 CCATGACAGGCGCGGCTTCGAGGAGGCGGACACCGGAGAGAGCTTTCGCGGCGCTC 359
QY 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140
DB 360 CCCTCGCGCTTACCGCGCTGTGCTACCACTCTCGGCGGACCAACCACTCCCGGACGAG 419
QY 141 LeuGluProLeuAlaTyrSerAspAspGlyValIleMetGlyLeuArgHisArgGlyLys 160
DB 420 CTCGAACCTTCGCTGCGGAGCGACAGCGCGTGTGATGAGCTTCGCGGACCGCGAGAG 479
QY 161 ProLeuTyrGlyValGluPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180
DB 480 CGCTGATGAGGCGTCCAGTTCACACCGAGGTGCATCGGACGACTTCGCGCGGAGATC 539
QY 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisAsnArgAlaArgArgHisGlyAlaAsp 200
DB 540 ATGGCCAACTTCGCACTTCGCTCCGCGCACACCGGCGAGTGTGACAGCGGCGCGAGC 599
QY 201 SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
DB 600 TGGGCTTACGAATCTCAAGTGGCGCGCTGACAGTCTCCGAGCCGAGAGAGTACCG 659
QY 221 ArgGlyCysLeuProGlyGluGlyThrThrPheTyrLeuAspSerSerSerValLeuGlu 240
DB 660 CCGGCTGCTGCGCGGCGGAGCGGCGGACGTTCTGCTGAGACGAGCTCCGCTCGAA 719
QY 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
```


D 720 GCGCGCTCCGCTCTCTCTCTCTGCGGACGACCGCGCGCTCCGCGAGTACCTCACC 779
Q TyrtarValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgThr 280
D 780 TACGCGCTCCGCGACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 839
Q 281 ArgArgProPhePheAsnTyrLeuGluGluGluGluGluGluGluGluGluGluGlu 300
D 840 GCGGCGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 899
Q 301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrLeuGly 320
D 900 CCGGACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 959
Q 321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
D 960 GCGGAGACCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 1019
Q 341 AlaAspArgAlaAlaAlaLeuAspHisGluGluGlyCysCysTyrLeuLeuAlaLeuAsp 360
D 1020 GCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1079
Q 361 ArgArgGlyHisAspAspArgValAlaArgAlaTrpLeuArgGluThrAlaGluThrThr 380
D 1080 CGCGCGGCGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1139
Q 381 GlyLeuAlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGlyTyrProGlu 400
D 1140 GCGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1199
Q 401 AlaAlaAlaGlyPheGlyProLeuAlaArgAlaHisAspAspAspAlaTyrLeuGly 420
D 1200 GCGGCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1250
Q 421 ArgTyrLeuAspGlyCysLeuGlyGluArgAsnGlyGluSerTyrGluTyrCysLeuThr 440
D 1251 -----TCGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1292
Q 441 AsnMetValThrAlaProThrThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuAla 460
D 1293 AACATGCTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1352
Q 461 IleSerProValProTyrGlyValAlaLeuLeuGluPheProGluLeuSerValLeuSerAla 480
D 1353 ATACAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1412
Q 481 SerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerTyrProIleGly 500
D 1413 TCGCCGAGCGGTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1472
Q 501 GlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAla 520
D 1473 GGGAGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1532
Q 521 GlyArgGluArgAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeu 540
D 1533 GCGCGGAGAGAGACCGCGCGCGAGAACCTATGATCTCTCTCTCTCTCTCTCTCTCTCT 1592
Q 541 AsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTyr 560
D 1593 AACAGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1652
Q 561 AlaProValHisGluLeuValSerThrIleArgGlyValArgLeuArgProGlyTyrSerThr 580
D 1653 GCGCGCGTGCACACGCTGTGTGCACATCCGGGAGAGCGGTGCGCGCGCGCGCACACG 1712
Q 581 AlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProGlyValArg 600
D 1713 GCGCGCGTGTACGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1772
Q 601 ThrMetGluTyrIleAspArgLeuGluGlyProArgGlyValTyrSerGlyAlaLeu 620
D 1773 CCATGAGATCATCAACCGCTGTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1832

Q 621 GlyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeu 640
D 1833 GATGCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1892
Q 641 AlaAspGlyGluAlaGluPheGlyValGlyValAlaIleValSerLeuSerAspGluGly 660
D 1893 GCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1952
Q 661 GluGluPheThrGluThrValValValAlaArgAlaMetValThrAlaLeuAspGlySer 680
D 1953 GAGAGAGTTCAGGACGACCGGTGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2012
Q 681 AlaValAlaGlyValArg 686
D 2013 GCAGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2030

RESULT 4
US-10-819-386A-1
; Sequence 1, Application US/10819386A
; Publication No. US20050089884A1
; GENERAL INFORMATION:
; APPLICANT: Korea Advanced Institute of Science and Technology
; APPLICANT: Shanghai Jiaotong University
; APPLICANT: LEE, Sang Yup
; APPLICANT: DENG, ZIXIN
; APPLICANT: CHEN, SHI
; APPLICANT: JEONG, Ki Jun
; APPLICANT: ZHOU, XIUFEN
; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polyketides
; FILE REFERENCE: P003-B015
; CURRENT APPLICATION NUMBER: US/10/819,386A
; CURRENT FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: KR10-2003-0074035
; PRIOR FILING DATE: 2003-10-23
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 138203
; TYPE: DNA
; ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1

Alignment Scores:
Pred. No.: 1,51e-163 Length: 138203
Score: 1766.00 Matches: 381
Percent Similarity: 63.93% Conservative: 71
Best Local Similarity: 49.43% Mismatches: 209
Query Match: 49.43% Indels: 46
DB: 21 Gaps: 13

US-10-089-514-2 (1-686) x US-10-819-386A-1 (1-138203)

Q 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGluTyrIle 20
D 18610 ATGCGACCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 18669
Q 21 GlyAlaAlaThrGlyGluProProValValAlaProAsn---AspAlaAspTyrSerArg 39
D 18670 TCCCGGCGCAAGCGCGCGGAGACCGAGTCAATCCGAAAGACGACCGCGCGCTGCG 18729
Q 40 LeuProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArg 59
D 18730 GGTCTCTCGACGCGGTTCGACAAACGTGTGCTCTCTCTCTCTCTCTCTCTCTCTCT 18789
Q 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGly 79
D 18790 CCGGCGGACTTGGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 18849
Q 80 ValGlyLeuGlyHisGluGlyIleAlaGluLeuPheGlyGlyThrValGlyLeuAlaPro 99
D 18850 GTCTGCTGGGCGACCAAGGCGCATGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 18909


```

OY      597  ||lyblybarkgrhrtwctgulllelleasparngleugluclnyproargylvaltlyrse 617
Db      1893  GAAACTGGCGACGAGTGGAGATCATCGACTCGCTGAGAGCCGAGGCCCGGGGGTGTACTC 1952
OY      617  rglvalaleuulyrtrpheaialeuSERGLYALALASpeuseerilevalilleagth 637
Db      1953  CGGAGCCATCGGCTACTTCCTCGGGTCCAGTGGGGGGCGGACCTGCCATCGCACGCCGAC 2012
OY      637  rtlevalleuAlaaspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSe 657
Db      2013  GGCGGTGTTCACCCACGCGGAGATGCACTTCGGCGCGGGCGGTGCGATCGCTCGGCTC 2072
OY      657  rASpGlnclnuclugluPheThrGluThnValValAlaAlaArgAlaMetValThrAla 677
Db      2073  CGATTCGGTCCGCGCGAGTACGACGAGATCTGTGAAGACGGCCGACCGATGCGGCTC- 2130
OY      677  uASpGlySerAlaValAlaGlyAlaArg 686
Db      2131  -CACCGGACCGGATCGCGGCTCGCGG 2157

RESULT 6
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272657
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Alignment Scores:
Pred. No.:      2,81e-144      Length:      9025608
Score:          1596.50      Matches:      354
Percent Similarity: 58.77%      Conservative: 75
Best Local Similarity: 48.49%      Mismatches: 247
Query Match:     44.68%      Indels:      55
DB:              15          Gaps:          12

US-10-089-514-2 (1-686) x US-10-156-761-1 (1-9025608)
OY      1  MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db      1470597  GTGAAGACCTCGCTCATCGACATTTACGACTCGTACACGTACACGTACCAACTCTTCCAGCTATC 1470538
OY      21  GlyGluAlaThrGlyGlnProProValValValProAsnAspAlaAspTrpSerArgLeu 40
Db      1470537  GCCAGAGTGGAACGCGCGAGACCGCGATGGTGGTCTCTCATATGACGCCGCCGCGAGACGATTT 1470478
OY      41  Pro---ValGluAspPheAspAlaIleValValSerProGlySerProGlySerProAspArg 59
Db      1470477  CCGAGTATCCGGGAGATTTCGCCAACAGTGATGTGTGTGCGCGGGGGCCGGACACCCCGCGAGAG 1470418

```

QY	60	GIuArghApPhegIyIIESeArArGAlAlIEthzrAspSerGlyLeuProValLeuGly	79
Db	1470417	AGAGGTGACTTCGGCATTCTGCGGAGAGTCTCCCGCGGCCCGCCGTCGGTGGC	1470358
QY	80	ValCysLeuGlyNIhsgInGlyIAlaGlnLeuPheGlyGlyThrValIGlyLeuAlaPro	99
Db	1470357	GTCCTGCTCGGCGACCAAGGAGCATTCGCGGAGGAGAGGGGCTGGGTGACCCCGGACCG	1470298
QY	100	GIuProMetNIhsgIArgValSerGlyValArgIshThrGlyIuAspValPheArgGly	119
Db	1470297	GAGCCCGCGCACCGGCATCTGTCCACCGTCCGCGCAACCGCCAGACCTGTTCGAGGGC	1470238
QY	120	LeuProSerProPheThrAlaValArgTyhIshSerLeuAlaAlaThrAsp--LeuPro	138
Db	1470237	CTGCCCCGAGAGTTTCAACCGCGGTCCGCTACACTCCCTGTGTGTAACGGACCGCTCCCA	1470178
QY	139	AspGluLeuGlyuPLeuLeuAlaTrpSerAspAspGlyValValMetGlyLeuArgNIhArg	158
Db	1470177	CTGAGCTCTGAGCCCAACCGCTGGCGGAGAGACGCGCTGTGAAGGACATCTGGCACCGC	1470118
QY	159	GIuTyPProLeuTyTrGlyValAlaGlnPheNIhProGluSerIIEgISeAspPheGlyArg	178
Db	1470117	TCCCGCGCGCTGTGGGGGTGCAATTCCACCCGGAAGTCCGCTTCACCAAGATTGGCCAC	1470058
QY	179	GIuIIEuMetAlaAsnPheArgAspLeuAlaLeuAlaNIhshArg-----	193
Db	1470057	CGAGATCTCTCGAATCTTCGCGAGATGACGGCTCCGCGCCCGCAGACTCTGCACGGAC	1469998
QY	194	---AlaArgNIhsgIyAla-----AspSerPro---	202
Db	1469997	AAGCGCGTACCTCCGGAACGGCGACACGAGCCCCCGTGGCGCCCGCTCCCAAGC	1469938
QY	202	-----	202
Db	1469937	GCCGAGGGGGGGGGGAGACACCCCGCTGTGTCCCGCGCCCGTCGGTGGCCCCC	1469878
QY	203	---TyTrGluLeuNIhsvAlaArgArgValAspValLeuProAspAlaGluGluVal-ArgAr	221
Db	1469877	TCCCAACCGCTTTCACACCCCGCGGATCGACCTGCGCGTGAACGCGCGAGGCGGCTTCAC	1469818
QY	221	GGIYCyLeuProGlyGlyNIhThrPheTrpLeuAspSerSerSerValLeuGluGly	241
Db	1469817	CGGATGTACACCGACGCGGCC-CCGCGCTTCTGCTGTGACAGCTCCCGGATCGAGAGGG	1469759
QY	241	YAlAsErArPheSerPheLeuGlyAspAspArgGlyProLeuNIhGlyTyTrIleuThrTy	261
Db	1469758	ACAAGTCCCGTTCCTTCGTTCTTCGGTACGCGACCGGACCGCTGGCCGAGTTTCGATCGGTA	1469699
QY	261	ArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrThrArgThrAr	281
Db	1469698	CGAGCTCGAGACGCGGCTCTGTGATGATCGAGCGGCCGCGGCGGCCCGTGCGCAAGTCAG	1469639
QY	281	GAArgProPhePheNIhTyTrLeuGluGluGlnIleuGluNIhArgArgValProValAlaPr	301
Db	1469638	GCCGACGCTCTTCACATACTCGAAACCGGACGTTGTGACCCGCTGAGGTGAT--GGCAC	1469582
QY	301	oGluLeuProPheGluPheAsnLeuGlyTyTrValGlyTyTrLeuGlyTyTrGluLeuVal	321
Db	1469581	GGGACTGCGGTTCACATTCACCGCGGGATATGTGGCTACTTCGCGCTACGAGGGAAGGC	1469522
QY	321	aGluThrThrGlyAspProAlaNIhArgSerProNIhProAspAlaAlaPheLeuPheAl	341
Db	1469521	CGAGTCGGGTCCTCGTCACACGTCATCGGGGCCGGAACCTCCGAGCGCTCGCGGCTTCGCG	1469462
QY	341	AspArgAlaIIEAlaLeuAspNIhsgInGlyGlyCysCysTyTrLeuLeuAlaLeu----	359
Db	1469461	CGACCGGCTGTGTCAGTGTGACATCGAGCGAGGACCACTACCCGCTGTCTCGCGCA	1469402
QY	360	---AspArgArgGlyNIhshAspArgGlyAlaArgAlaTrpLeuNIhArgNIhArgNIh	378
Db	1469401	GAAACACCGGAGGCGCGACGACGACGAGCGCC--GACTGTGCTCAACGCGCGATGCGCCA	1469345
QY	378	rIleuThrGlyLeuAlaValArgAlaPro--AlaGluProThrProAlaMetValPheGly	397

```

Db 1469344 GCTCAGTTGTGTCCTCGGCAAGCCGACCGCCCGCCGCGCCAGC----- 1469296
Qy 397 YLLEPGLUALAIAAIAAGLYPHEGLYPROLEUALAARGIAAARGHISAPLYSAPAL 417
Db 1469295 -GCACCCCGCTCGATCCCGGAGCCGCTGCTGTGCTCCG-----CACCGTAGAC 1469246
Qy 417 ATYRLLEULYARGILEASPLGUCYLEULYSGIULIARAGASNDLYGLUERTYRGUUL 437
Db 1469245 CTACCTCGCGACATCGGACCTCGGACCTGCAAGCGGAATCAAGACGCGACCGATTACGAGAT 1469186
Qy 437 ECYLEUTHRASMMEVALTHRALAPROTHGLUALATHRALALEUPROLEUTHYRSEAL 457
Db 1469185 CTGTCTACCAACCGCCGCGGACGCTGCCCGCCCGCTACAGACCCCTACGACTTACCGGGT 1469126
Qy 457 ALEUARGALALESERPROVALPROLYRGYALALEULEUGLUPHEPROGLUEUSERVA 477
Db 1469125 ACTGCGCTGCTCGACCGCGCGGCTGACCGGCTTATCTGCGGGTTCGTGACCTCGATGT 1469066
Qy 477 LLEUSERLASERPROGLUARGPHELEUTHRIIEGLYALASPGLYGLYVALGLUSERLY 497
Db 1469065 GCGCGGCTGCTCGCCCGGAGGTTTCTGCGGATCACCGGACGCGCTCGCGGAGCGAG 1469006
Qy 497 SPROILELYSGLYTHRARPPROARGGLYGLYTHRALAGLUGLUSPGLUARGLEUARGAL 517
Db 1469005 ACCCTGTAAAGGACCGCGCGCCCGCGCGGACCGCGGAGAGAGACCCCGCTACCGGA 1468946
Qy 517 AASPYLEUALAGLYARGGLULYASPARAGALAGLUSLEUMELILEVALASPLEUALAR 537
Db 1468945 CGCGCTCACCGACGACGACAAAGCCCGCGGAGAACCTGGTAGTGTCTGACCTGCTCCG 1468886
Qy 537 GAENASPLEUASMERVALCYBALALEGLYSERVALHISVALPROARGLEUPHEGLIUA 557
Db 1468885 CAACGACCTGGCGCGGCTGCGCGGACCGGAGCTGAAGTCAACCGCTGATGCGGAC 1468826
Qy 557 LGIUTHRYALAPROVALHISGLINEUVALSERTHRIIEARGLYARGHEUARGPROGL 577
Db 1468825 CGAGACCTGCGCCACCGCTGACCAAGTTGTTCTCCACGCTGAGAGGCGGCTCGCGAGGG 1468766
Qy 577 YTHRSERTHRALAALCYSEVALARGALAAAPHEPROGLYGLYSERMERTHRGYALAPR 597
Db 1468765 CARGGAGCGGTGAGACTGCGTAAGTGTCTGCTTCCCGGAGGTTCCGTGACCGGCGGCC 1468706
Qy 597 OLYSLYARGTHMERGLUILEIIEASPARGLEUGLUGLYPROARGLYVALYLYSE 617
Db 1468705 GAAGAATCGGACGATGAGATCATCGACTGCTGAGACCGAGCGCGCGGCTGTACTC 1468646
Qy 617 RGLYALALEUGLYTRYPHEALALEUSERGLYALAAIASPLEUSERLILEVALIIEARGTH 637
Db 1468645 CGGAGCCATCGGCTACTTCCGAGTGCAGTGGCGCGGCGGCGGCTCCGCTCCGACCGCCAC 1468586
Qy 637 RILEVALLEUALAASPGLYGINIAGLUPHEGLYVALIGLYVALALEVALSERLEUSE 657
Db 1468585 GCGCGTGTTCACCGACGCGGAGATGCACTCCGCGCGCGGCGGCTGCTCCCTCGGCTC 1468526
Qy 657 TASPGLINGLUGLUPHERTHRGUHTHRVALIIEVALIIEVALIIEVALIIEVALIIE 677
Db 1468525 CGATCCGCTCGGAGATGACGACGATGCTGTGAGACCGCGCACCGGATGCGCGCC-- 1468468
Qy 677 UAEPGLYSERLALVALIAGLYALARG 686
Db 1468467 -CACCGGACCGGATCGCGGCTGCGG 1468441

```

```

; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, MAKINO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1111
; LENGTH: 1860
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
; US-09-738-626-1111
Alignment Scores:
Pred. No.: 4,25e-138 Length: 1860
Score: 1467.00 Matches: 328
Percent Similarity: 60.06 Conserved: 78
Best Local Similarity: 48.52 Mismatches: 204
Query Match: 41.62 Indels: 66
Gaps: 13
US-10-089-514-2 (1-686) x US-09-738-626-1111 (1-1860)
Qy 1 MetcargThreuleuulleaspaenlyraspsrphethrhisaneuphegintytle 20
Db 1 ATGCGGGTTTATATTATGATATATGATCTTTCACGTTTATATCCGACCTATGTC 60
Qy 21 GYGLUALATHRGYGLINPROPROVALVALIIPROASPARALASPTPSEARGLEU 40
Db 61 GAAGAGTTTACGGGTACGACGACCTGTGTGTCCTTAATGATCAAAAATAGATGATG 120
Qy 41 PROVALGLUASPHASPARALILEVALISERPROGLYPROGLYSERPROASPARGLU 60
Db 121 CTT-----TTGACCGCGGATCTCTCACTGCGCGGCGGACCGCGGCTTGC 171
Qy 61 ARGASPHAGLYLISERARGALALIEHRAASERGLYEUUPROVALLEUGLYVAL 80
Db 172 GCTGATTTTGTATCTGTGCGAGCGCTTGAACGCTGCAAGCTTCCGATTTGGGTGTC 231
Qy 81 CYLEUGLYHISGLNGLYILEAGLNULEUPHEGLYGLYTHRYALIGLYLEUALAPROGLU 100
Db 232 TGTTTAGGCACAGAGGATGCGTGGCGCTTATGCGGATGATTTGGCGCCAG 291
Qy 101 PROMENHISGLYARGYALSERGLUVALARGHISTHCHLYGLUASPVALPHEARGLYEU 120
Db 292 CCGGTCCAGGTGAGGTTTCCGAGTCCAGTCCAGTATGTTTATTTGACGAGTATC 351
Qy 121 PROSERPROPHETHRALVALARGYRHSERLEUALALATHRASPLEUPROASPARGLU 140
Db 352 CTGGAAGCTTACAGCGCGGCTGCTTATCATCTGATGATGAGCATTTGACATGAGTCTT 471
Qy 141 LEUGLUPROLEUALATPSEASPARAGLYVALIIEVALIIEVALIIEVALIIEVALIIE 160
Db 412 TTGAAGCTTACAGCTACAGCGATGATGTTGATGATGATGATGATGATGATGATGATG 471
Qy 161 PROLEUTHRGYVALGINPHEHISPROGLUSERLIEGLYSEASPHAGLYARGGLUIE 180
Db 472 CCGAGTGGGCTGTGAATTTATCCGGAATTTATTTGGAATTTGGAATTTGGAATTTGGA 531
Qy 181 METALASPHASPARALALEUALIIEHISARGALARGARGHISGLYALASPAR 200
Db 532 ATTAAGAACTTCTTATTTATGCGGACATATCC----- 567

```

OY	201	SerProTyrGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg	220
Db	568	-----TGGCAACTCAGCAGAAAACTATTCGCGCTCACCGTTGATTACGACGCGGTTTTT	621
OY	221	ArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSerValLeuGlu	240
Db	622	GAACATTCTTGGCCATTCTCCCATGCTTTTGGCTCGATTGAT-----	666
OY	241	GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr	260
Db	667	---GCCCAAGGAACAGCTATCTTGATGTCAGCGGTCTCTCGACGCAAAAAACC	723
OY	261	TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr	280
Db	724	CATATGTCGGGAGGGG-----	741
OY	281	ArgArgProPhePheAsnTyrLeuGluGluGluLeuAlaArgArgValProValAla	300
Db	742	-----GATTTCTTCACTGCTGAAGAAGAGATCTCGCCCAAC-----TCAGTTGCG	789
OY	301	ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLys	320
Db	790	CCGGGTCAA-----GTTTCGTCTTGCGCGGTGGTTAGTTAGCTTAAGCTTAA	843
OY	321	AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe	340
Db	844	GCGGAAGCTGGCGCACGGCTGGCGCACCTTCGATCTTCGGATGCCACCTCATTTTT	903
OY	341	AlaAspArgAlaIleAlaLeuAspHisGlnGluGlyCysCysTyrLeuLeuAlaLeuAsp	360
Db	904	GCCGATCGCGCATGCGATGCGATGGAATCGGATCAG-----GTTCCGTTCTGCGTGG--	954
OY	361	ArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGluThrAlaGluThrLeuThr	380
Db	955	-----GGGAGACAGGAC-----GAGTGGTTGAAGAAACATCAAGACTGCAT	999
OY	381	GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyTyrLeuPro	399
Db	1000	AATCTTGTGCGCCCGCGGATCTGCTGCC-----	1029
OY	400	GluAlaIleAlaGlyPheGlyProLeuAla---ArgAlaArgHisAspLysAspAlaTyr	418
Db	1030	-----GACACCTCGCTTTCGAGGTTCCAGATTCCAAAAGATGAGAT	1071
OY	419	LeuLysArgGlyLeuAspGluCysLeuLysGluIleArgAsnGlyGlyLysSerTyrGluLys	438
Db	1072	CTCGCAAAATTCGAGAGCCGAGAGCGTGAATTCTCGCGCGAATCGTAAGAATCTGC	1131
OY	439	LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu	458
Db	1132	CTGACCACAAACTTCAGGGCACCACTGATGCGCCCTCTGGCTGCTTACTTGACACTG	1191
OY	459	ArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeu	478
Db	1192	CGTGGGCGCAATCCACCGCAATGTCGTGATCTTCAGCTGGGGGGAATCCTCATTTTG	1251
OY	479	SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro	498
Db	1252	AGTTCCTCGCGGAGCGGTTCATCATCATTGATTCGGAGGGTATGTGGATCAAAGCC	1311
OY	499	IleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAsp	518
Db	1312	ATTAAGGACACAGGCGCGTGGGGAACAACGCGCAAGAAACCAAGAAATCATTTCTGAG	1371
OY	519	LeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn	538
Db	1372	CTGGCGCAATATCCTTAAAGATGTGTCAAAAACTTGATGATCGGATTGTGTCGCAAC	1431
OY	539	AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu	558
Db	1432	GACTTAGCCCGCGCGCTTGGCCACCACTGAAGTAAACATCAAGCTTTTGAGCTGCAG	1491
OY	559	ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr	578

```

Db      1492  ACCAAGCCAGCAGCTCCACCACTTGAGCACCCTCTCTCCAGAGTTGGGGCCCA--CGC 1548
Oy      579    SerThrAlaAlaCySeVAlArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Db      1549  AGTCGGATGTAGTGTGCGTCGCGGCGACGACTTCCCGGATGTTGATGATGACTGGTCCCCANAAG 1608
Oy      599    LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyrSerGly 618
Db      1609  CTGGCGACCATGAGATCATCATGATGACGTGGAGGCACGCTCCTCGCGGATATTACTCAGGT 1668
Oy      619  AlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Db      1669  GCGTTGGGATATTTTTCCTCCGACGGGCGCAGTTGATCTCCACAGGATCAGAACTCTC 1728
Oy      659  ValIleuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAsp 658
Db      1729  GTCATCCAGAAACAATCACGTGAGTACGAGATGGGCGGTGACACTTCTTGCTGTGTGAT 1788
Oy      659  GlnGluGluGluIupherhrgIurThvalValIlysAlaArgAlaMetVal 674
Db      1789  CCGAGGCGTAGTGCGAGAGAAATCCGGGTTAAATCAAGGCGCTGTGCTG 1836

RESULT 8
US-10-494-675-25
/ Sequence 25, Application US/10494675
/ Publication No. US20050019877A1
/ GENERAL INFORMATION:
/ APPLICANT: Zelder, Oskar
/ APPLICANT: Pompejus, Markus
/ APPLICANT: Schroder, Hartwig
/ APPLICANT: Kroege, Burkhard
/ APPLICANT: Klopprogge, Corinna
/ APPLICANT: Haberhauser, Gregor
/ TITLE OF INVENTION: Genes coding for metabolic pathway proteins
/ FILE REFERENCE: BGI-163US
/ CURRENT APPLICATION NUMBER: US/10/494, 675
/ CURRENT FILING DATE: 2004-05-04
/ PRIOR APPLICATION NUMBER: PCT/EP02/12141
/ PRIOR FILING DATE: 2002-10-31
/ PRIOR APPLICATION NUMBER: DE 101 54 292.1
/ PRIOR FILING DATE: 2001-11-05
/ NUMBER OF SEQ ID NOS: 164
/ SEQ ID NO 25
/ LENGTH: 2005
/ TYPE: DNA
/ ORGANISM: Corynebacterium glutamicum
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (101)..(1975)
/ OTHER INFORMATION: RXA00579
US-10-494-675-25

Alignment Scores:
Pred. No.:      4,69e-138      Length:      2005
Score:          1487.00      Matches:      328
Percent Similarity: 60.06%      Conservative: 78
Best Local Similarity: 48.52%      Mismatches:   204
Query Match:     41.62%      Indels:       66
Db:              21          Gaps:         13

US-10-089-514-2 (1-686) x US-10-494-675-25 (1-2005)
Oy      1    MetArgThrIleuLeuIleAspAsnTyrAspSerPheThrIleAsnLeuPheGlnTyrIle 20
Db      116  ATGGCGCGTTTAATTAATTAATGAATTAATGATTTTTCACGTTTAATCTCGGCACCTTAGTG 175
Oy      21  GlyGluAlaThrGlyGlnProProValValAlaProAsnAspAlaAspTyrSerArgLeu 40
Db      176  GAAAGAGGTACGGGTCAAGCACCTGTGTGTGCTTATATATCAAGAAATAGATGAGATG 235
Oy      41  ProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArgGlu 60

```


Db 236 CTT-----TTGACGCCGTCATCTCTACCTGCGCCGGGACGCGCGGCTTCCG 286
Qy ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
Db 287 GCGATTTTGGTATCTGTGCGAGGGCGCTTACGCGTGCACGCCGTTCCGATTTTGGGTGTG 346
Qy CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
Db 347 TGTTTAGGCCACACGAGGCGATTCGTTGGCCGTGATGTTGATTTGGCCGCCAGG 406
Qy 101 PromethisGlyArgValSerGlnValArgHisThrGlyGlyAspValPheArgGlyLeu 120
Db 407 CCGGATCCACGCGTGAAGTTTGCAGATCCCATGATGATTCAGGTTTATTTGACGGGATC 466
Qy 121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140
Db 467 CTTGAAACGTTTGAAGCGGCGTCTTATCATCTGATGGCGACGCCGCTTCCGGAGTCA 526
Qy 141 LeuGlnProLeuAlaIlePheSerAspAspGlyValValMetGlyLeuArgHisArgGlyLys 160
Db 527 TTGAAAGCTACAGCTTACGAGGATGATGTTGATCATGGCATTTGGCACATGAAGTCTT 586
Qy 161 ProLeuTrpGlyValGlnPheHisProGlnSerIleGlySerAspPheGlyArgGlyIle 180
Db 587 CCGCAGTGGGGTGTGCAATTTTCATCCGGAATCTATTGTGGACAACTTTCGCGCATCAGATC 646
Qy 181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleHisArgAlaArgAlaGlyValAsp 200
Db 647 ATTAAAGACTCTCTTAATTTTACCGCGCAATATGCG-----682
Qy 201 SerProTyrGlnLeuHisValArgArgValAspValLeuProAspAlaGlnGluValArg 220
Db 683 -----TGGCAACTCAGCGAGAAACATATTCGCTCCGCTGATTTAGCAGCGCGTCTTT 736
Qy 221 ArgGlyCysLeuProGlyGlnGlyThrThrPheTrpLeuAspSerSerValLeuGlu 240
Db 737 GAAACATCTTTGCCCATCTCCATCCATGCTTTTGGCTCGATGAT-----781
Qy 241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGlyTyrLeuThr 260
Db 782 ---GCCAAGGAAACAGCTATCTTGTGATGCCAGCGTCTCTCGACCGACACAAAC 838
Qy 261 TyrArgValAlaAspGlyValIleSerValArgGlySerAspGlyThrThrArgThr 280
Db 839 CATTAATCTCGCGCAGGG-----856
Qy 281 ArgArgProPhePheAsnTyrLeuGlnGlnLeuGlnArgArgValProValAla 300
Db 857 -----GATTTCTTCACTGCTAAGAGGATCTGCCGCCAAC-----TCACTTGGC 904
Qy 301 ProGlnLeuProPheGlnPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGlyLeuLys 320
Db 905 CCGCGTCA-----GGTTTCTGCTTGGCTGGGTTGTATGCTGTATGAGCTTAA 958
Qy 321 AlaGlnThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db 959 GCGGAAGCTGCGGACCGGCTGCGCACACTTCCGATGCGGCACTTCAATTTT 1018
Qy 341 AlaAspArgAlaIleAlaLeuAspHisGlnGlyCysArgTyrLeuLeuAlaLeuAsp 360
Db 1019 GCGGATCGCGCATCGAGTGAATCGGATCAG-----GTTCGCTTGGTGGCGTTC-- 1069
Qy 361 ArgArgGlyHisAspAspArgGlyAlaArgAlaIleTrpLeuArgGlnThrAlaGlnThrLeuThr 380
Db 1070 -----GGGAGGACGAGAC-----GAGTGGTTTGAAAGAAACATCAGAGAGCTGCAT 1114
Qy 381 GlyLeu-----AlaValArgAlaProAlaGlnProThrProAlaMetValPheGlyIlePro 399
Db 1115 AATCTTGTCTCGCGCGGATACCTGCTCC-----1144
Qy 400 GluAlaAlaIleArgPheGlyProLeuAla-----ArgAlaArgHisAspLeuAspAlaTyr 418
Db 1145 -----GGACACCTCTGCTTTCAGAGTTGAGATTCGAAAGATGAGTAT 1186

Qy 419 LeuLysArgIleAspGluCysLeuLysGlnIleArgAsnGlyGlnSerTyrGlnIleCys 438
Db 1187 CTGCAGAAATTTGCGAGAGCGCCAGAGCTGATTAATCTCGCGCAATCTGTAATAATCTGC 1246
Qy 439 LeuThrAsnMetValThrAlaProThrGlnAlaThrAlaLeuProLeuTyrSerAlaLeu 458
Db 1247 CTGACCAAAATCTCAGGCGCACCACTGATGTGGCCCTCTGGCTGCTTATCAGACTG 1306
Qy 459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGlnPhePheProGlnLeuSerValLeu 478
Db 1307 CGTGGGCGCAATCCACCGCATATGCGCTATCTTACGTGGGAGATACCTTAATTTTG 1366
Qy 479 SerAlaSerProGlnArgPheLeuThrIleGlyAlaAspGlyGlyValGlnSerLysPro 498
Db 1367 AGTTCCTCGCGGAGCGGCTTATCATCATGATTTAGCCAGGAGTATGTGAATCAAGCCC 1426
Qy 499 IleLysGlyThrArgProArgGlyGlyThrAlaGlnGlnAspGlnArgLeuArgAlaAsp 518
Db 1427 ATTAAAGGACCAAGCGCGGTGGGGAACAGCGCAAGAGCAAGAAATCATTTGCTGAG 1486
Qy 519 LeuAlaGlyArgGlyLysAspArgAlaGlnLeuMetIleValAspLeuValArgAsn 538
Db 1487 CTGCGAGTAATCTTAAGATCGTGCAAAACTTGATGATCGTGATTTGGTCCCAAC 1546
Qy 539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGlnValGlu 558
Db 1547 GACTTACCGCGCGCGCTTGGCCACCAAGTAAACATTCACGTTTTCAGCTGCA 1606
Qy 559 ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
Db 1607 ACCTAGCCACAGTCCACCACTTGTACGACCGCTCTGCGAGATTTGGGCGCA---CGC 1663
Qy 579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Db 1664 AGTCCAGTGAAGTGTGCGCGGAGCATTCGCGGTGTGATGATCTGGTCCCAAG 1723
Qy 599 LysArgThrMetGlnIleIleAspArgLeuGlnGlyProArgGlyValTyrSerGly 618
Db 1724 CTGCGACCAATGAGATCATCATGATGAGCTGGAGAGCTCTCGCGTATTTTACAGGT 1783
Qy 619 AlaLeuGlyThrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
Db 1784 GCGTTGGATATTTTCCCTCCAGCGCGAGTGAATCTCTCCATGATGATCAGAACTCTC 1843
Qy 639 ValLeuAlaAspGlyGlnAlaGlnPheGlyValGlyAlaIleValSerLeuSerAsp 658
Db 1844 GTCATCCAGAAACATCAGCGGAGTACGAGATGGCGGCGTCACTTGTGCTGTGAT 1903
Qy 659 GlnGlnGlnGlnPheThrGlnThrValValLysValAlaArgAlaMetVal 674
Db 1904 CCGAGGCGTGAAGTGGAGGAAATCCCGTTAAATCAGCGGCTCTGCTG 1951

RESULT 9
US-09-738-626-1
; Sequence 1, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCES: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484

```
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patent ver. 3.0
; SEQ ID NO 1
; LENGTH: 3309400
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1

Alignment Scores:
Pred. No.:          6,64e-134          Length:          3309400
Score:              1487.00            Matches:          328
Percent Similarity: 60.06%             Conservative:    78
Best Local Similarity: 48.52%           Mismatches:     204
Query Match:        41.62%             Indels:         66
                                           Gaps:          13

US-10-089-514-2 (1-686) x US-09-738-626-1 (1-3309400)

QY      1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlyTyrIle 20
Db      1052021 ATGCGCGTTTAAATTATGATTAATTATGATTCCTTTCACGTTTATCTCGCCACCTATGTC 1052080

QY      21 GlyGluAlaThrGlyGlnProProValValProAsnAspAlaAspTyrSerArgGlu 40
Db      1052081 GAAGAGGTTACGGGTCAGGACCTCTGTGTGCTCTATATATCAAGAAATAGATGACATG 1052140

QY      41 ProValGluAspPheAspAlaIleValSerProGlyProGlySerProAspArgGlu 60
Db      1052141 CTT-----TTGCGACGCGCATCTCTCACCTGCGCGGCGCACGCGCGGTTGCG 1052191

QY      61 ArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGlyVal 80
Db      1052192 GCTGATTTTGGTATCTGTGACGCGTCATTTAGCGTGACGCGTTCGATTTTGGGTGTG 1052251

QY      81 CysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaProGlu 100
Db      1052252 TGTATTAGCCACCAAGGACATGCGTTGGCTATAGCGCGATGTTGATTTGGCGCCAGG 1052311

QY      101 ProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGlyLeu 120
Db      1052312 CCGGTCACGCGTGAAGTTTGCAGATACACCATGATGTTCAAGTTTATTGGACGCGATC 1052371

QY      121 ProSerProPheThrAlaValArgTyrHisSerLeuAlaIleThrAspLeuProAspGlu 140
Db      1052372 CCTGAACGTTTGAAGCGGTCGCTTATCATCTCATGCTGCGCAACCCGCTTCCGCGAGTCA 1052431

QY      141 LeuGluProLeuLeuIleTyrSerAspAspArgValValMetGlyLeuArgHisArgGluLeu 160
Db      1052432 TTGAAAGCTACACGTACACGACGATGATGTTAATCATTCGATTTGGACATGAAAGTCCTT 1052491

QY      161 ProLeuTyrPylValGlnPheHisProGluSerIleGlySerAspPheGlyArgGluIle 180
Db      1052492 CCGCAGTGGGCGTGGCATTTCAATCCGATCTAATTGGTGGACAATTCGCGCCATCAGATC 1052551

QY      181 MetAlaAsnPheArgAspLeuAlaLeuAlaHisIleArgAlaArgArgHisGlyAlaAsp 200
Db      1052552 ATTAAAGAACTTCCCTAATTATGAGCGCGCACATATCGC----- 1052587

QY      201 SerProGlyGluLeuHisValArgArgValAspValLeuProAspAlaGluGluValArg 220
Db      1052588 -----TGGCACTCTACGGAAGAAACTATTCCTCCCTACGCTTAATTCAGACGCGCTTTT 1052641

QY      221 ArgGlyCysLeuProGlyGluGlyThrThrPheThrLeuAspSerSerSerValLeuGlu 240
Db      1052642 GAAGAACTTCTTGGCCATTCCTCCCATGCTTTTGGCTCATGAT----- 1052686

QY      241 GlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAlaGluTyrLeuThr 260
Db      1052686
```

```
Db      1052687 ---GCCCAAGAAACAGACTATCTTGGTATGCCAGCGGTCTCTCGCACGCACAAAAACC 1052743

QY      261 TyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThrThrArgThr 280
Db      1052744 CATAAATCGCGCAGGCG-----

QY      281 ArgArgProPhePheAsnTyrLeuGluGlnLeuGluArgArgValProValAla 300
Db      1052762 -----GATTTCTTCACTGGCTAAAGAGAGATCTCGCGCCAAAC-----TCAGTTGGCG 1052809

QY      301 ProGluLeuProPheGluPheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLeu 320
Db      1052810 CCGGTCGCA-----GTTTTCCTTGGCTGGCTGGTTGATGCTTAAAGCTTTAA 1052863

QY      321 AlaGluThrThrGlyAspProAlaHisArgSerProHisProAspAlaAlaPheLeuPhe 340
Db      1052864 GCGGAAGCTGGCGCACGGGCTGGCGCACACTTGCAGTCTTCGCGATCGCACCTCATTTT 1052923

QY      341 AlaAspArgAlaIleAlaLeuAspHisGlnGlyCysCysTyrLeuLeuAlaLeuAsp 360
Db      1052924 GCCGATCGCGCCATCGCAGTGGAAATCGATCAG-----GTTCCGTTGCTGGCGTTG--- 1052974

QY      361 ArgArgGlyHisAspAspGlyAlaArgAlaTyrPheLeuArgGluThrAlaGluThrLeuThr 380
Db      1052975 -----GGGAGACAGGAC-----GAGTGGTTTGAAGAAACCATCAAGAGCTGCAT 1053019

QY      381 GlyLeu---AlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIlePro 399
Db      1053020 AATCTTGTGCGCCCGCGGATACCTGCGTCC----- 1053049

QY      400 GluAlaAlaAlaGlyPheGlyProLeuAla---ArgAlaArgHisAspIysAspAlaTyr 418
Db      1053050 -----GGACACCTCGCTTTCAGATTCGAGATTCGAAAGATGAGTAT 1053091

QY      419 LeuLysArgIleAspGluCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCys 438
Db      1053092 CTTGCACAAATTCGACAGACCCAGAGCTGATTAATCTCGCGGCAACGATGAATCTCG 1053151

QY      439 LeuThrAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeu 458
Db      1053152 CTGACCAAAACTTGAAGGCGACACTGATGTGCGCCCTGTGCTCTATCTAGACATG 1053211

QY      459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGluLeuSerValLeu 478
Db      1053212 CGTGGGCGCAATCCACCGCATATGCTGTATCTTCACTGGGGAATCACTTATTTTG 1053271

QY      479 SerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerLysPro 498
Db      1053272 AGTTCTCGCGGAGCGGTTCAATCACTTGAATTCGCGAGGATGTGGAATCAAAAGCCC 1053331

QY      499 IleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArgAlaAsp 518
Db      1053332 ATTAAAGGACCAAGCGCGGTGGCGACAGCGCAAGAAAGCAACMAAATCATTTGCTGAG 1053391

QY      519 LeuAlaGlyArgGlyLysAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsn 538
Db      1053392 CTGCGAGTAATCTTAATAATGTGCGAAAACCTGATGATGTGGAATTTGGTCCGCAAC 1053451

QY      539 AspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
Db      1053452 GACTTACCGCGCGCGCTTTGGCCACCAAGTTAAACATCCAAAGCTTTTCGACGTGCA 1053511

QY      559 ThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
Db      1053512 ACCTAGGCAAGTCCACCAACTTTCAGACACCGTCTTCGCAAGTGGGGCCA---CCG 1053568

QY      579 SerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLys 598
Db      1053569 AGTCGATTTGAAGTGTGCGCGAGCATTCGCCGAGTGTTCATGACTGTGCTGCCCAAG 1053628

QY      599 LysArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyValTyrSerGly 618
Db      1053629 CTCGCAACATGAGATCATCATGATGAGCTGAGGAGAGCTCTCGCGGTATTACTCAGGT 1053688
```



```
QY 455 TysSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPheProGlu 474
Db 1318 TACAAACACTTGGCCAGGAAAGTCCCGGTCACTACTGCTTTTGGCTTTAGATGAT 1377
QY 475 LeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyVal 494
Db 1378 TTTCAAGTCTTTCCTTCCTTCATGAGCGGTTTATTAAGTCCACAAGACCGGGTGTG 1437
QY 495 GluSerProIleTyrGlyThrArgProArgGlyGlyThrAlaGluIleAspGluArg 514
Db 1438 ACCACTAAGCCCTCAAGGGGACGTCACAGGGGACAGATACCAAGAGACCAAGAC 1497
QY 515 LeuArgAlaAspLeuAlaGlyArgGluTyrAspArgAlaGlyAsnLeuMetIleValAsp 534
Db 1498 TTGATGAAAGGCTCCGACGTAAAGAAACCAAGGACGAAACCTGATGATGTCAG 1557
QY 535 LeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeu 554
Db 1558 CTCTCCGCAATGATCTGGGCGTTTGTGAATCGGGTGGGTGAGTACCAAACTG 1617
QY 555 PheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeu 574
Db 1618 ATGATGTAGAGCTTACCTACCTCCACAGCTGATTACAGGTTTCAGGCGGATC 1677
QY 575 ArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThr 594
Db 1678 AAGAAGACCTAGATGTTGTTGAAGTGAAGAAACACCTCCGGGCGGTCATGACT 1737
QY 595 GlyAlaProGlyAlaArgThrMetGluIleIleAspArgLeuGluGluGluProArgGly 614
Db 1738 GGGGCCCTTAAAAAAGAACCCCTGGAATATTGATGACTTGAACAGATGCCAAGG 1797
QY 615 ValTyrSerGlyAlaLeuGluTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleVal 634
Db 1798 ATGATATCTGGGCTATGCTTTTAAAGCAATTCACATTAACCTTCAACATCGTC 1857
QY 635 IleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIleVal 654
Db 1858 ATCCGACCAAGCTGTGTGGAAGACAGACAGCAAGCATGCGCTGGGGGGTGCATTTGTC 1917
QY 655 SerLeuSerAspGlnGluGluPheThrGluThrValValIleValAlaArgAlaMetVal 674
Db 1918 ATGCTATCTGATCTCTGAAGAGAGTTGATGAAGTTTAAAGCTAAGGGGCGCTTG 1977
QY 675 ThrAlaLeu 677
Db 1978 TCCGCTTG 1986

RESULT 11
US-10-501-282-6651/C
; Sequence 6651, Application US/10501282
; Publication No. US20050203280A1
; GENERAL INFORMATION:
; APPLICANT: MCWICHAEL, JOHN CALHOUN
; APPLICANT: ZAGURSKY, ROBERT JOHN
; APPLICANT: RUSSELL, DAVID PARRISH
; APPLICANT: FLETCHER, LEAH DIANE
; TITLE OF INVENTION: ALDOTOCCUS OTITIDIS OPEN READING FRAMES (ORFS) ENCODING
; TITLE OF INVENTION: POLYPEPTIDE ANTIGENS, IMMUNOGENIC COMPOSITIONS AND USES THEREOF
; FILE REFERENCE: AM100780 I2
; CURRENT APPLICATION NUMBER: US/10/501,282
; PRIOR APPLICATION NUMBER: 60/333,777
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: 60/426,742
; PRIOR FILING DATE: 2002-11-18
; PRIOR APPLICATION NUMBER: PCT/US02/36123
; PRIOR FILING DATE: 2002-11-25
; NUMBER OF SEQ ID NOS: 6653
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 6651
; LENGTH: 1754382
; TYPE: DNA
```

```
; ORGANISM: Alloiococcus otitidis
US-10-501-282-6651

Alignment Scores:
Pred. No.: 1,41e-120 Length: 1754382
Score: 1350.00 Matches: 296
Percent Similarity: 58.71% Conservative: 105
Best Local Similarity: 43.34% Mismatches: 248
Query Match: 37,78% Indels: 34
DB: 22 Gaps: 12

US-10-089-514-2 (1-686) x US-10-501-282-6651 (1-1754382)

QY 1 MetArgThrLeuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGlnTyrIle 20
Db 7807 ATGAAGTCACTCATTTTCAATTCATTAATCTTATCTTACTTACACTTATTAACGCTTATT 7748
QY 21 GlyGluAlaThrGlyGlnProProValValProAsnAsp--AlaAspTrpSerArg 39
Db 7747 GGGAGGTCAACAGGAAAGGCCCATGTGATATAAAATGACCAATGATCAACAAAGAA 7688
QY 40 LeuProValGluAspPheAspAlaIleValValSerProGlyProGlySerProAspArg 59
Db 7687 CTATTGATCTCGACTTTGATATGATCATATTATTTCAACCGGCGCTGTGTGTCGACCG 7628
QY 60 GluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeuGly 79
Db 7627 GACAGGACTTTCGGCTTGTGCGGACAGTAAATTGAAAAGCTTGACAAAGCTTATCTTTGGC 7568
QY 80 ValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAlaPro 99
Db 7567 ATTTCCTGGGCGACAGGCGCATTTATTATTAAACGGTGGGCACTTATGAGGGCGCTGAT 7508
QY 100 GluProMetHisGlyArgValSerGluValArgHisThrGlyGluAspValPheArgGly 119
Db 7507 ATCCCATCATGTCGCGGACAGATTCGCTCAATTAAGGGAATAATATCTTTGACGCGC 7448
QY 120 LeuProSerProPheThrAlaValArgTyrHisSerLeuAlaAlaThrAspLeuProAsp 139
Db 7447 ATAGACCAAGGCTTTGAACTCAACGCTACCTCATTTGTTGTAAGAC-----AAG 7394
QY 140 GluLeuGluProLeu-----AlaTrpSerAspAspGlyValValMetGlyLeuArg 156
Db 7393 GAATTAGAAATATTAATTCATTCATGACCAAAACGATGATGAATTGTCAATGGCCCTGTC 7334
QY 157 HisArgGluLeuPheLeuTrpGlyValGlnPheHisProGluSerIleGlySerAspPhe 176
Db 7333 CACAAGACCAAGCCCATCTATGCGGTCCCAATTCACCCGAAATCCATTCGCCACCCGATAT 7274
QY 177 GlyArgGluIleMetAlaAsnPheArgAspLeuAlaLeuAlaHisArgAlaArgArg 196
Db 7273 GGGAAAAAATATGATGAAAACCTTATGCTTGTGTCAGACTACTAT----- 7226
QY 197 HisGlyAlaAspSerProTyrGluLeuHisValArgArgValAspValLeuProAspAla 216
Db 7225 -----AACCAATCCAGCCTTTATTATGAAAAGTTCCGGGAACATGAGACACAC 7178
QY 217 GluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTrpLeuAspSerSer 236
Db 7177 CAAGACCTTACGATCAATTTGCGCAAAATTTGATGACCAAGTACTTTGGCTCAGCTCAGC 7118
QY 237 SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256
Db 7117 AAGGTGAGCCTGGCTGTGCACGGTTTCCATCTTTGGATGGCTGGACCAAGCGGGCG 7058
QY 257 GluTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspGlyThr 276
Db 7057 CACACTTAAATATGATGATCCACCAAAAGAGTGAAGAAAACAAAGATGACCAATCCGCA 6998
QY 277 ThrThrArgThr--ArgArgProPhePheAsnTyrLeuGluGluGluLeuGluArgArg 295
Db 6997 GCAGTTGAAGGCTCCAGACTGACATCTTTCTTACTTA-----AAGGCCAACGACCA 6944
```

296 ARGVALProVALIAProGluLeuProPheGluPheAsnLeuGlyTYrValGlyTYrLeu 315
 6943 AAATGGCTTATGAGCCGGCGCTTGGCTTTCAGCTTACAGCTTAAGGCTAATGGCTAATATC 6884
 316 GJTYRGJLeuLYaAGJUTHrThrgly--AsPProLahIaArgSerProhIsPPro 334
 6883 GGCTATGAGGTCAAAAAGATACAGTTCAAGTCCAAACAAAGACCACTCAACCTTATCG 6824
 335 AsPAlaIaPheLeuPheAlaAsPArgAlaIeAlaLeuAsPHisGlnGlnGlyCysCys 354
 6823 GATGCTCACTTACTTACTTATGTGTACCGCGGCTTGCTGTATGACACACAGAAGAGGACCTT 6764
 355 TYrLeuLeuAlaLeuAsPArgHrArgGlyHisAsPAsPArgJAlaArgAlaITrPLeuArgJLu 374
 6763 TACTTCCTCTCTTAC-----CAGGTGAC-----CAGATTTGATTTCAACG 6722
 375 ThrIaGJUTHrLeuThrgJlyLeuAlaValaArgAlaProLahGJUPrOTHrProLahMet 394
 6721 GTCAAAAGACAGGCTCAAG---CAGCGATGCAAAAGTTGACGAAAGAGA----- 6677
 395 ValPheGlyIleProGlnAlaIaIaAGJYPheGlyProLeuAlaArgAlaArgHisAsP 414
 6676 -----CCAAGAGGCAAGCAAGACTCCACAGCATGGCTTGCTGTCAAA---CAC 6632
 415 LysAsPAlaTYrLeuLYaArgJLeAsPArgJCySeuLYaGJUIleArgAsnGJGJLuser 434
 6631 AAAGCAGGCTATATATAAAGACATTTAAACCATCAAGACTGTATCAAGCTGGGGAAAGT 6572
 435 TYrGJUIleCYeLeuThrAsmMetValThrAlaProThrgJAlaIaThrAlaLeuProLeu 454
 6571 TATGAAGTTGCTTACCAACCGCGCTGATATTAGGGGTCAAGATTGATGCTTGCCTTAT 6512
 455 TYrSerAlaLeuArgAlaIleSerProValProTYrGJAlaLeuLeuGluPheProGJLu 474
 6511 TACAACAACCTTGGCGCAGAAAGTCCCGGTGAGTACTGCTGCTTTTGCCTTTAGATGAT 6452
 475 LeuSerValIleuSerAlaSerProGJLuArgPheLeuThrIleGJAlaAsPArgJLyAl 494
 6451 TTTCAGTTCTTCTCTTCCATGAGGCGTTTATTAATGTCGACMAAGACCGGGTGTGTG 6392
 495 GJuserLYaProIleLYaGJLYThrArgProArgGJLyJUTHrAlaGJLuAsPArgJLuArg 514
 6391 ACCACTTAAGCCCATCAAGGGAGGCTCAGACGGGGACAGATGACCAAGAAAGACCAAGAC 6332
 515 LeuArgAlaAsPLeuAlaGJLYaArgJLuLYaAsPArgJAlaGJLuAsnLeuMetIleValAsP 534
 6331 TTGATAGAAGGCTCCGCGCAGTGAAGAAACCAAGGACGAAGAACTGATATTGTGCAC 6272
 535 LeuValIaArgAsnAsPLeuAsnSerValLYaAlaIleGJYSerValHisValProArgJLeu 554
 6271 CTCTTCCCAATGATCTGGCGCGTTTGTGAAATCGGGGTGGTTGAAGTACCAGAACTG 6212
 555 PheGJLuValGJUTHrTYrAlaProValHisGJLeuValSerThrIleArgGJLYaArgJLeu 574
 6211 ATGAGTGAAGACCTACTCTCCACCTTCACACAGCTGTTCACACAGGTTTCAGCGCGGCTC 6152
 575 ArgProGJLYThrSerThrAlaIaLYaCYaSerAlaArgAlaIaPheProGJLYIleSerMetThr 594
 6151 AAAGAAAGACTAGAGTGTGGTTGAAGTCTTAAGAAACCACTTCCCGGGGGGTTCATGACT 6092
 595 GJLYaLProlYLYaArgThrMetGJUIleIleAsPArgLeuGlnGJLYaProArgJLy 614
 6091 GGGGCGGCTTAAAAAAGAACCTCGAATTTATTGATGACTTGAAGCAGTTCCCAAGAGGG 6032
 615 ValTYrSerGJLYaLeuGJLYThrPheAlaLeuSerGJLYaIaIaAsPLeuSerIleVal 634
 6031 ATCTATTCGGGACTAATGGCTTTTATTCACCAACTTCCACTATAGACTTCAACATGCTC 5972
 635 IleArgThrIleValLeuAlaAsPArgJLYaIaGJLuAsnGJLYaIleGJLYaIleVal 654
 5971 ATCCGGAACAGTGTGTGGAAAGACAAAGACAGACATCGGCTGGGGGGGTGCATTTGTC 5912
 655 SerLeuSerAsPArgJLuGlnGJLuPheThrgJUTHrValValLYaIaIaArgAlaMetVal 674

[illegible]

```

Db      2525 GGTGCTACCTCTTAAATCAATCCATCCAGGTAAATAAAGCTTCAAGTAGTGGCG 2466
QY      129 TyrHisSerLeuAla-----AlaThrAspLeuProAspGluLeuAlaIleTyr 146
Db      2465 TATCACTCACTTGTATAGAACAGACTCTATCTGAGAGATCTTATATCAATAGCATG 2406
QY      147 -----SerAsp----- 148
Db      2405 ACTGCTTCCAAAAATGCTCTCATTTCTTGAAGAAGCATAGCTGATATTAAGTACAGT 2346
QY      148 ----- 148
Db      2345 ACCCTGTGGGGATCATTTGATTAATTATTCGTAAACAACAGTCAAGTGCAGTACCAGT 2286
QY      149 -----AspGly-----ValIleMet 153
Db      2285 GATGAAAAATGCCCAAGCATAAACGATGCAAGTGAATGATGCTACAGGGTTCATG 2226
QY      154 GlyLeuArgHisArgGluLysProLeuTyrGlyValGlnPheHisProGluSerIleGly 173
Db      2225 GGGGTTAGGCACTCTACAGAGCCCTCATGTGAGTCAAGTTTCACCCCGAGAGTGTGCT 2166
QY      174 SerAspPheGlyArgGluIleMetAlaSerPheArgAspLeuAla----- 188
Db      2165 ACTCATTTATGGAAGACAGATATTTCAAAACCTTCAGAAGATAAACAACGACTTGGATTA 2106
QY      189 -----LeuAla 190
Db      2105 CAGACACCATGCTTCAGGAAAAAAGATCATGTTCCAGATCTCAAGAACTTGTGTGCA 2046
QY      191 His-----HisArgAlaArg----- 195
Db      2045 AATGACTTGTATCATTTCTGCAAGAGTGAACCTTGGGATTCGTGGCCCTTGTGCTT 1986
QY      196 -----ArgHisGlyAlaAspSerProTyrGluLeuHisValArgArgValAsp----- 211
Db      1985 CCAAGCGAAGCAGTGGGGCAAAATGCTTAAGGTTCATGAGAAAAAGATGTAACTTC 1926
QY      212 -----ValLeu-----ProAspAla 216
Db      1925 CTCATTCGCAATAGTGGCTCTGAAAAATTTCAGTGTCTTTGGCCATCATACCGCT 1866
QY      217 GluGluValArgArgGlyCysLeuProGlyGluGlyThrThrPheTyrLeuAspSer 236
Db      1865 GAAGAT-----ACATTTGGCTGGATTCCTCA 1839
QY      237 SerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProLeuAla 256
Db      1838 TCAGTTGACCGAATAGGGCAGCATTTTCATCATGGCGGCAAGGGCGGCTTTGG 1779
QY      257 GlyTyrLeuThrTyrArgValAlaAsp-----GlyValAlaSerVal 270
Db      1778 AAGCAATATGACATTTCACTTCGCAAGTCAACAGCAATTTGTGAGAGAACTTACTAT 1719
QY      271 ArgGlySerAspGlyThrThrThrArgThr-----ArgArgProPhePheAsnTyrLeu 288
Db      1718 CGAGATGCTTATGTTGATCTGCAAGAACTTCTCAAGAGATGTTCTTGGATTTCCCT 1659
QY      289 GluGluGlnLeuGluArg--ArgArgValProValAlaProGlu--LeuProPheGlu 306
Db      1658 GACAAAGAGATGCAATCCATTCATTAATCAATTGAAGAGATTAAGACACTTCATTGTGAC 1599
QY      307 PheAsnLeuGlyTyrValGlyTyrLeuGlyTyrGluLeuLysAlaGlu--ThrThrGly 325
Db      1598 TTCATGCTGCGATTTGTTGGATTCATAGATATGCTTTAAAGTTAAATGCGATGCAATCA 1539
QY      326 AspProAlaHisArgSerProHisProAspAlaAlaPheLeuPheAlaAspArgAlaIle 345
Db      1538 TCTAATATGTCAAATATCAAGTACCCCTGATGCTGCTTCTTCTGTATTAACCTAGTT 1479
QY      346 AlaLeuAspHisGlnGluGlyCysTyrTyrLeuLeuAlaLeuAsp-----ArgArg 362

```

```

Db      1478 GTGGTTGATCAACAACATGGGAGTGTATCAATTTATATCATATCATGATGAATATTTCTCT 1419
QY      363 GlyHisAspAspGly-----AlaArgAlaTyrPheArgGluThrAla 376
Db      1418 GGTAAATGAGATGGAGATTACCAAAAACATCAATACATAGTTATGTTAGTACAAATACTGAG 1359
QY      377 GluThrLeuThrGlyLeuAlaValArgAlaPro----- 387
Db      1358 AAGAAAGCTTCTCAGATGATGATCCATGGGCCCAAGATTATCGATCAATGAAAACTGTGCG 1299
QY      388 -----AlaGluProThrProAlaMetValPheGlyIlePro 399
Db      1298 ATCAATGGGAATCATTTATCCATATCATATCCAGTGTGAATAGCAAAAGATTTGTCAAG 1239
QY      400 GluAlaAlaAlaGlyPheGlyProLeuAlaArgAlaArgHisAspLysAspAlaTyrLeu 419
Db      1238 AATATCA-----AAAGATGAATATATAC 1218
QY      420 LysArgIleAspGluCysLeuLysGluIleArgAsnGlyGluSerTyrGluIleCysLeu 439
Db      1217 AAGAGATGTCAGAGATGCTCGATTTACATAAAGACGAGAAAGCTATGAATTGTGCTTA 1158
QY      440 ThrAsnMetValThrAlaProThrGlu--AlaThrAlaLeuProLeuTyrSerAlaLeu 458
Db      1157 ACTACTCAGATGAAGAAAGAACGGATTAATATGATGCTTTGAACCTTAACCTGAAATTTG 1098
QY      459 ArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPhe-----ProGluLeuSer 476
Db      1097 CGAAAAACAATAATCCAGGCCCTTATATGAGCTTGGCTTAATCTTCCTCAAAAAACCTGAT 1038
QY      477 ValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyValGluSer 496
Db      1037 AATATGTTGCTTTCCTGTA----- 1017
QY      497 LysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeuArg 516
Db      1017 ----- 1017
QY      517 AlaAspLeuAlaGlyArgGluLysAspArgAlaGluAsnLeuMetIleValAspLeuVal 536
Db      1016 -----AGTAAAAAAGATCAAGCTGAGAACTTGTATGATATTTGTGACCTCCCTTA 972
QY      537 ArgAsnAspLeuAsnSerValCysAlaIleIleGlySerValHisValProArgLeuPheGlu 556
Db      971 AGAAACGATCTTGTAAGGTTGTGAACCTGGGAGTGTACATGTTCTCGCTCATAGAT 912
QY      557 ValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgPro 576
Db      911 GTGGAACTCAATTAACCTGTATACATCCATGATGAATCAATTCGTGGAACCAAAATGTGCG 852
QY      577 GlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAla 596
Db      851 GACCTAAGCCCTGTGATTTGTAAAGCTGCTCTTCCAGAGGTTCAATGACCGGGGCC 792
QY      597 ProLysLysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTyr 616
Db      791 CGAAAGTTATGATTCATGAGATCTTGTATCACTTGAAGAAATGATGCCGAGAGAAATATAC 732
QY      617 SerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArg 636
Db      731 TCAGATGCTGTTGATCTTTTCATATTAACAAGACTTTGACTGTAAATTCGTAGACAA 672
QY      637 ThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSerLeu 656
Db      671 ACGATTTGCTTCACATATGAGAGAGCTTGATTTGGGCGAGGGGCTATTTGTAGCATTTG 612
QY      657 SerAspGlnGluGluPheThrArgGluThrValValLysValAlaArgAla 672
Db      611 TCAGATCCAGAAAGCAGATGAATGATGATGCTGTTAAAGCAAAAGCT 564

```

RESULT 13
 US-10-282-122A-17825
 ; Sequence 17825, Application US/10282122A


```
Db      1084 GCCGGGAAACATATGAACTCTGACCAACCACTATTCGCCGAAGTACACGGTGAG 1143
Qy      451 AlaleuPro-----leuTyrSerAlaleuArgAlaleuSerProValProTyrGlyAla 468
Db      1144 GTCCATTCGGCGGCAATGATCAAGCCGTAGACACCGCACTACTGCGCTTATGCGCTT 1203
Qy      469 leuLeuGluPheProGluLeuSerValleuSerAlaSerProGluArgPheLeuThrIle 488
Db      1204 CTGGTTGTCAACGATGATGAGCAGTGAATTTCCGCGCTCACCAAGAGATTTATCACATG 1263
Qy      489 GlyAlaAspGlyValGluSerLysProIleLeuGlyThrArgProArgGlyGlyThr 508
Db      1264 AAC---GACCGCATGTGTGTCTCTCCATCCATCAAGAAACCAAGAAACCCACGCCGAC 1320
Qy      509 AlagluGluAspGluArgLeuArgAlaAspLeuAlaglyArgLysAspArgAlaglu 528
Db      1321 CGGGAAGAAAGACCGCGCTTGCGCGACGATCTGGCACCAACCCCAAGACCGCGCGGAA 1380
Qy      529 AsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySer 548
Db      1381 AACCTATGATCGTCCAGCTTGTTGGCGAAGACCTGCGCGCTTCCGATCAGCGACG 1440
Qy      549 ValHisValProArgLeuPheGluValGluThrTyrAlaProValHisGluLeuValSer 568
Db      1441 GTCCGTGTGCACAACATATGTGCATGCACACGCTTACACCGTGACACGATGATTTCT 1500
Qy      569 ThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPhe 588
Db      1501 ACCGTGAAAGGGAACCTTCCGCCCAACGATGCTTATGCACTCTCCACCGCACCTTTC 1560
Qy      589 ProGlyGlySerMetThrGlyAlaProLysLysArgThrMetGluIleIleAspArgLeu 608
Db      1561 CCAAGCGGCTTCCATGATGAGTGGGGCGCCCAACCGCACCTGACCTCACTACCGAGCTC 1620
Qy      609 GluGluGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAla 628
Db      1621 GAAGGCAAAACAACGCGGGTTTACTCCGATGATGATGATATGCGGACGATCTCCGC 1680
Qy      629 AlaAspLeuSerIleValIleArgThrIleValleuAlaAspGlyGlnAlaGluPheGly 648
Db      1681 ACCGATTTCTGATGATGATCCGACCGCTGCTCACACCCACCACTTGAGCTTATGGG 1740
Qy      649 ValGlyValIleValleuValSerLeuSerAspGlnGluGluPheThrGluThrValVal 668
Db      1741 GTGAGAGGTGCATCATCGGCTTTTCAGACCCCGCAAGATGGGAGAAATCACACACA 1800
Qy      669 LysAlaArgAlaMetValThrAlaLeu 677
Db      1801 AATATCCCGGTGCTTCTCGACCTACTC 1827

RESULT 14
US-10-282-122A-16977
/ Sequence 16977, Application US/10282122A
/ Publication No. US20040029129A1
/ GENERAL INFORMATION:
/ APPLICANT: Wang, Liangsu
/ APPLICANT: Zamudio, Carlos
/ APPLICANT: Malone, Cheryl
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl
/ APPLICANT: Zyskind, Judith
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John
/ APPLICANT: Carr, Grant
/ APPLICANT: Yamamoto, Robert
/ APPLICANT: Forsyth, R.
/ APPLICANT: Xu, H.
/ TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
/ FILE REFERENCE: EITRUA.034A
/ CURRENT APPLICATION NUMBER: US/10/282.122A
/ PRIOR FILING DATE: 2003-02-20
/ PRIOR APPLICATION NUMBER: 60/191,078
```

```
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/230,335
/ PRIOR FILING DATE: 2000-09-06
/ PRIOR APPLICATION NUMBER: 60/230,347
/ PRIOR FILING DATE: 2000-09-09
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/267,636
/ PRIOR FILING DATE: 2001-02-09
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 78614
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 16977
/ LENGTH: 1906
/ TYPE: DNA
/ ORGANISM: Clostridium difficile
US-10-282-122A-16977

Alignment Scores:
Pred. No.: 1,51e-89 Length: 1906
Score: 1001.00 Matches: 253
Percent Similarity: 52.52% Conservative: 112
Best Local Similarity: 36.40% Mismatches: 245
Query Match: 28,02% Indels: 85
DB: 17 Gaps: 20

US-10-089-514-2 (1-686) x US-10-282-122A-16977 (1-1906)
Qy      4 leuLeuIleAspAsnTyrAspSerPheThrHisAsnLeuPheGluTyrIleGlyAla 23
Db      4 CTATATGATGACATATATATATATCTTTGTATACAACTGATCCAAATATATA--GAAAGAA 60
Qy      24 ThrGlyGluProProValValAlaProAsnAspAlaAspTyrSerArgLeuProValGlu 43
Db      61 TTAGGGAAGAAAGTTGTATGATAAAGAAAT-----AATGAATCAAAATAAAGT 108
Qy      44 AspPhe-----AspAlaIleValSerProGlyProGlySerProAsp 58
Db      109 GACATTGAAGAACTAAATCCAGAGGTATTTGTATCTCCAGACCGGTCTTCCA--- 165
Qy      59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThr-----AspSerGlyLeuPro 76
Db      166 -----AAAGAAAGCTGGTATTTGATGATATTTGTAGACATTTTAAGGGAAAAACCT 219
Qy      77 ValLeuGlyValCysLeuGlyHisGlnGlyIleAlaGlnLeuPheGlyGlyThrValGly 96
Db      220 ATACTAGAAATTTGTTTAGGTGACCAACTATTTGACATGTTTGGAGAGATATTATA 279
Qy      97 LeuAlaProGluProMetHisGlyArgValSerGluValArgHisThrGlyGluSerVal 116
Db      280 AAAAGCAACAAGCGCTGTGATGAGAAAGATATAGTATCAACATCTAATAAAGAGATT 339
Qy      117 PheArgGlyLeuProSerProPheThrAlaValArgTyrHisSerLeuAla-----Ala 134
Db      340 TTTAGGGACCTTAAATAATCTTTAAATGTCACCTAGATATCATCTAATAATTAATGATTCC 399
Qy      135 ThrAspLeuProAspGluLeuGluProLeuAlaTyrSerAspAspGlyValValMetGly 154
Db      400 AATACAGTTCCAAAAGAAATTAAGAAATTAACAGCAATTAACGTAAGAGGAAATATAGGG 459
Qy      155 LeuArgHisArgGlyLysProLeuTyrGlyValGlnPheHisProGluSerIleGlySer 174
Db      460 ATTAGGATTAAGAAATATTAATTAATAGAGAGTACAGTTTCATCCAGAACTATTTGTGCA 519
```

```

Qy 175 AspbheglYargLuiIeMeAlaAspPheargAspleuAlaIeAlaHisIeArgAla 194
    ::::::::::::::::::::
Db 520 GAGATAGTCATGAAATGCTTAAATTTT-----ATPAACAAGAGCTTAGAGAA 567
Qy 195 ArgArgHisIeGlyAlaAspSerProTgIuleuHis-ValArgArgValAspValLeuPr 214
    ::::::::::::::::::::
Db 568 AGGTATACAT-----GTGTAATATAGTAAAGAGATTAATTAATTAAT 609
Qy 214 aaAspAlaGluGluValArgArgGlyCyLeuProGlyGluGlyThrThrPheTPrLeuAs 234
    ::::::::::::::::::::
Db 610 AAATCTCTTGAAGATATTACT--ATTTTGAAGATGACATGACGTTTATATTAAGA 666
Qy 234 pSerSerValLeuGluGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyPr 254
    ::::::::::::::::::::
Db 667 CAGTCGCTATGATTAAGAAATTTGGAGAGATTTCTTTATTAAGTACGCCA-----CC 720
Qy 254 OLeuAlaGluThrLeuThrThrArgValAlaAspGlyValValSerValArgGlySerAs 274
    ::::::::::::::::::::
Db 721 ATTT--AGGTATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 751
Qy 274 pGlyThrThrThrArgArgArgArgProPhePheAsnTyIeGluGluGluLeuGluArg 294
    ::::::::::::::::::::
Db 752 -----GMAAAATCCT--TTAGAAATTTTAAAGAAAGAACTACATTA 789
Qy 294 gArgArgValProValAlaProGluLeuProPheGluPheAsnLeuGlyTyValGlyTy 314
    ::::::::::::::::::::
Db 790 GTATAGGCTGTAAATGATTAATTAATCTTCCTTTGTAGGA-----GGTGTGTGGGATA 843
Qy 314 rLeuGlyTyGlyIuleuLeuValGluThrThrGlyAspProAlaHisArgSerProHis- 333
    ::::::::::::::::::::
Db 844 TTTATCTTATGATTTA-----GTAATTACATGAAAAATTTACTACAGAC 888
Qy 334 -----ProAspAlaAlaPheLeuPheAlaAspArgAlaIeAl 346
    ::::::::::::::::::::
Db 889 TGCTGTAGATGATTTGAGATGCTGATATGATTTGATTTGATTAATCATGTAATAGT 948
Qy 346 alaAspHisIeGluGluGlyCySerTyIeLeu-----LeuAlaLeuAs 360
    ::::::::::::::::::::
Db 949 TATAGACCATCTTGTGCAAAAACTTATATTAAGTACTCCAAACATGATATTCGATTAAGA 1008
Qy 360 pArgArgGlyHisAspAspGlyAlaArgAlaTrpLeuArgGlyThrAlaGluThrLeuTh 380
    ::::::::::::::::::::
Db 1009 AGAAAGATTAATGATTAATTAAGAACAAAGAAATTTAAAGAAAGAAAGAAAGTATAGA 1068
Qy 380 rGlyLeuAlaValArgAlaProAlaGluProThrProAlaMetValPheGlyIleProG 400
    ::::::::::::::::::::
Db 1069 TAGCATATGTTATGA--GMAAAAGAAATTAACACTATTAAGACTT----- 1111
Qy 400 uAlaAlaIaIaGlyPheGlyProLeuAlaArgAlaArgHisAspLeuAspAlaTyIeLeu 420
    ::::::::::::::::::::
Db 1112 -----AAATCTAATTTTAAACAAGAAATTTTAAAA 1143
Qy 420 sArgIleAspGlyCyLeuLeuGluIleArgAsnGlyGluSerTyGlyIleCyLeuTh 440
    ::::::::::::::::::::
Db 1144 TGCAGTTCAAAGCTTGAAGAGTACATTAAGCAAGAGATATATATACCTAATTAAC 1203
Qy 440 rAsnMetValThrAlaProThrGluAlaThrAlaLeuProLeuTySerAlaLeuArgAl 460
    ::::::::::::::::::::
Db 1204 ACAAAAGCTTCAGTCGACAAACAGAACTTACAGCTTGAATGATAGGACTTAAGAG 1263
Qy 460 aIleSerProValProTyGlyAlaLeuLeuGluPheProGluLeuSerValLeuSerAl 480
    ::::::::::::::::::::
Db 1264 ATTTAGCTCGGCAACATTTGGGCACTTAACTTTGAAGATGACATATATATTAACAA 1323
Qy 480 aSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyValGluSerTySProIleTy 500
    ::::::::::::::::::::
Db 1324 TTCTCCGAAGAGATTTTAAGATGT--GTTAATTAAGAGATTAAGAACTGACCAATTA 1380
Qy 500 sGlyThrArgProArgGlyGlyIleThrAlaGluGluLeuAspGluArgLeuArgAlaAspLeuAl 520
    ::::::::::::::::::::
Db 1381 AGGTACTGCTCAAGAGAGAAAGATTAAGAAAGATTAAGAACTTAACACAGAACTTAG 1440

```

```

Qy 520 aGlyArgGlyLeuAspArgAlaGluLeuMetIleValAspLeuValArgAsnAspLe 540
    ::::::::::::::::::::
Db 1441 AAATAGTCAAAAAGATAGAGCTGAGTACTTATGATAGTATTAAGATTAAGAAATGATAT 1500
Qy 540 uAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGluThrTy 560
    ::::::::::::::::::::
Db 1501 TGGCAGAAATATCTAAACAGTGAAGTGTAAAGTTCCAGACACTATTTGATTAAGACATA 1560
Qy 560 rAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerTh 580
    ::::::::::::::::::::
Db 1561 TGCAAATGTAAACACTTACTTCCAGAGTGTAGAGTGAAGTAAAGATTAAGATGAC 1620
Qy 580 rAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProValSer 600
    ::::::::::::::::::::
Db 1621 TACTGATGTATTAAGAACTTTCCAGGGGGGTCTATTAACAGAGACACTTAAGATTAAG 1680
Qy 600 gThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValTySerGlyAlaLe 620
    ::::::::::::::::::::
Db 1681 AGCAATGCAAAATATATAGCAATTAAGCCAAAGCAAGAAAGATTAATATGCTGTTCAAT 1740
Qy 620 uGlyTyTrpPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLe 640
    ::::::::::::::::::::
Db 1741 AGGTATATAGTGTATTAATGATGATATGATTTTAATATATGCAATTAAGCAATTAATCA 1800
Qy 640 uAlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnG 660
    ::::::::::::::::::::
Db 1801 AAATGACAAAAAGCTTACTTCAAGTTGAGAGAAATTAATCTTGGATTTGACCCAGA 1860
Qy 660 uGluGluPheThrGlyThrValValIleValAlaAspAlaMetVal 674
    ::::::::::::::::::::
Db 1861 TGAAGATATACAGAAACACTTATTAAGTAAATTAATCAATATAG 1903

RESULT 15
US-10-398-221-9/c
; Sequence 9, Application US/10398221
; Publication No. US20040018514A1
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 9
; LENGTH: 684707
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-9

Alignment Scores:
Pred. No.: 1,43e-69 Length: 684707
Score: 834.00 Matches: 234
Percent Similarity: 45.14% Conservative: 82
Best Local Similarity: 33.43% Mismatches: 224
Query Match: 23.34% Indels: 160
DB: 17 Gaps: 16

US-10-089-514-2 (1-686) x US-10-398-221-9 (1-684707)
Qy 4 LeuLeuIleAspAsnTyAspSerPheThrHisAsnLeuPheGlnTyIleGlyAla 23
    ::::::::::::::::::::
Db 375647 TTAATTAATGATCATTAATGATCATTTATTAATTAATTTTATTAAGAGCTT 375588

```

QY 24 ThGlyGlnProProValValProAsnAspAlaAspTyrSerArgLeuProValGlu 43
Db 375587 AAGGAAGATGTCGGGAGTTCCTGCTAGAGAT-----TTTCGCTCA 375543
QY 44 AspPheAsp-----AlaIleValValSerProGlyProGlySerProAsp 58
Db 375542 GCATTCGACCAATTAGCAGCTGAAAGGTATTCTTCGCGACCTGGATTCGCGAT 375483
QY 59 ArgGluArgAspPheGlyIleSerArgArgAlaIleThrAspSerGlyLeuProValLeu 78
Db 375482 -----GATTTCCAGAAAGTTTAGCTCTCTAGACAAATTACGCTCCCATTTCTC 375432
QY 79 GlyValCysLeuGlyYhiSgInGlyIleAlaGlnLeuPheGlyGlyThrValGlyLeuAla 98
Db 375431 GGGATTGGCTTCGGCCATCAAAAGATTGGTCATTTTTCGGGCAAAAGTTGTTACAGCG 375372
QY 99 ProGluProMetHISGlyArgValSerGluValArgHISThrGlyGlyAspValPheArg 118
Db 375371 GACGTTCTGTTACGGAAAACTAGCATATCTCCGATCAGCGTGAAGGATTATTTGCC 375312
QY 119 GlyLeuProSerProPheThrAlaValArgTyrHISerLeuAla-----AlaThrAsp 136
Db 375311 GCACTTGATTCGAGATTCTCAGTTCGCCGCTACCATTTCTTAGTCATTCGCCAATACT 375252
QY 137 LeuProAspGlyLeuGlnProLeuAlaTyrPheSerAspAspGlyValValMetGlyLeuArg 156
Db 375251 GTCCCACTGAATTGAAGTACGCGCTGTAAACAGATGCGGTTTAAATGGGTTAGAG 375192
QY 157 HIsArgGlyLeuProLeuTyrGlyValGlnPheHISerGlySerArgPhe 176
Db 375191 CATCTTAACAAAGCGATTACAGTTCATTCATTCCTTAACCAATCTTCTCGAAMAC 375132
QY 177 GlyArgGlyIleMetAlaAsnPheArgAspLeuAlaLeuIleHISArgAlaArgArg 196
Db 375131 GGAACGCAATCTTAGAAAACCTT----- 375108
QY 197 HIsGlyAlaAspSerProTyrGlyLeuHISValArgArgValAspValLeuProAspAla 216
Db 375108 ----- 375108
QY 217 GluGluValArgArgGlyCys-LeuProGlyGlnGlyThrThr-PheTyrPheAspSers 236
Db 375107 -----ATACGACTGGGAGAGCTGCCAAATGAGTTTATTACGTTTGAATTTTGAAGGTT 375054
QY 236 er---SerValLeuGlnGlyAlaSerArgPheSerPheLeuGlyAspAspArgGlyProL 255
Db 375053 CGGTAAAGATTTTGAAAAACCAATTCGCACTGGGCGACCGCAT-----T 375006
QY 255 euAlaGlyTyrLeuThrTyrArgValAlaAspGlyValValSerValArgGlySerAspG 275
Db 375005 TATCGGAAGTACTT----- 374992
QY 275 TyrThrThrThrArgThrArgArgProPhePheAsnTyrLeuGlnGlnLeuGlnArgA 295
Db 374991 -----CTATATTAAGAAAGCGCGCGGAAGAAAGCCCAAAATTCAG 374955
QY 295 rGArgValProValAlaProGlyLeuProPheGlnPheAsnLeuGlyTyrVal---GlyT 314
Db 374954 GAAA-----TACGTGCTCGTT 374937
QY 314 TyrLeuGlyTyrGlyLeuLysAlaGlnThrThrGlyAspProAlaHISArgSerProHISr 334
Db 374936 TTGTGATTATAG-----GCCCTCTCTGCTTTTCGTAAG----- 374902
QY 334 roAspAlaAlaPheLeuPheAlaAspArgAlaIleAlaLeuAspHISGlnGlnGlyCysC 354
Db 374902 ----- 374902
QY 354 ySTyrLeuLeuAlaLeuAspArgArgGlyYhiSAspAspGlyAlaArgAlaTyrLeuArgG 374
Db 374902 ----- 374902
QY 374 IuThrAlaGluThrLeuThrGlyLeuAlaValArgAlaProAlaGluProThrProAlaM 394

Db 374901 -----AATTTAAAAACCAACACCTCGCGAAGCATGCCACTTA 374862
QY 394 erValPheGlyLe-----ProGluAlaAlaIleArgPheGlyP 407
Db 374861 TCTGGTTTGATGATACGATTAACCTTCAGACTGCACCTGAACGAAACTAGACACTGCAC 374802
QY 407 roLeuAla---ArgAlaArgHISAspLysAspAlaTyrLeuLysArgIleAspGlyCysL 426
Db 374801 CACTTCTTTTAAATGATAGGATTTACCCGAATTAATCGGAAAAAATAGCAGATTA 374742
QY 426 euLysGlyIleArgAsnGlyGlySerTyrGlyIleCysLeuThrAsnMetValThrAla- 445
Db 374741 AAGCAAGAAATTCCTACTCGGAAACACATACCAATCAACTACCTGTTCCGCTCCAAAGTG 374682
QY 446 -----ProThrGluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerPro 464
Db 374681 AGGTACCAAGTGAAGTTTTCGCCCAAGCTTCTTTGAATCGCTGCACAAAGTTGGAAAG 374622
QY 464 aLProTyrGlyAlaLeuLeuGlnPheProGlyLeuSerValLeuSerAlaSerProGlu 484
Db 374621 CAACATATACAGCGCTACTGGAAGCGATGAATTCAGATTAATCTCCGCTTACCCGAGAC 374562
QY 484 rGPheLeuThrIleGlyAlaAspGlyGlyValGlnSerLysProIleLysGlyThrArgP 504
Db 374561 TATTTTAAAA---TGGAAAGAAACATCTTGACACAGCTCCAAATGAAGCACCATTC 374505
QY 504 roArgGlyGlyThrAlaGlnGluArgGluArgGluArgLysArgLysArgLysArgGluL 524
Db 374504 GCCGCGAAGTACGGAACAGCTGACTTGAAAGCACATGATTTGTTAAAAATGATTTCTA 374445
QY 524 ySAspArgAlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValC 544
Db 374444 AAATCGTCTGTAATAATGATGATTTGTTGATTTACTAGAAATAGACCTTGGGGTGAAT 374385
QY 544 yehAlaIleGlySerValHISValProArgLeuPheGlyValGlnThrTyrAlaProValH 564
Db 374384 CGGTGCTGTAGCGTAAGAGTCCCTAAATTAATGACTTTAGAACCTTACCCACAGTGT 374325
QY 564 IeGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaIaCysV 584
Db 374324 GGCAAATGACGCTTACTGTACCGCAAAACACCTCTGTATCTAGTTTAACCGCAGTTT 374265
QY 584 aLArgAlaAlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrMetGluI 604
Db 374264 TCGAGCGCTTTTCCCTTGTGTTCAATACTGGCGCCCAAAAGCCCGACAAATGGAA 374205
QY 604 IeIleAspArgLeuGlnGlnGlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheA 624
Db 374204 TTATTTCAAGCTAGAAAGATTCACTCGAGCGGTATACGTGGCGGATTTGGATTCCTAG 374145
QY 624 IeLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyG 644
Db 374144 AACCTAATGAAATGGGATTTTAAATGTCCAAATTCGAACGATTCATTCAGATGAATA 374085
QY 644 IuAlaGlnPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGlnGluPheT 664
Db 374084 AAGCAACTATAGTGTGTGTCGCGATGTCTTGGATTTCTGATGCCGAGACGAATTTT 374025
QY 664 hArgIuThrValValLysAlaArgAlaMetValThrAlaLeuAspGlySerAlaVal 682
Db 374024 CCGAGATTGATGCAAAATCAGCCATTTTGAAGAAAGCAACCAAGTTTCTTTAAAT 373969

Search completed: October 6, 2005, 00:28:12
Job time : 20:08.7 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 7504.63 Seconds
(without alignments)
3479.464 Million cell updates/sec

Title: US-10-089-514-2
Perfect score: 3573
Sequence: 1 MRTLIDNYSFTNHLFOYI.....VVKRAMVTALDGSAAVAGAR 686

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q/cg2_1/USFTO_spool_h/US10089514/runat_04102005_105744_8043/app_query.fasta_1.1621
-DB=EST -QFMT=fastcap -SUFFIX=ret -MINMATCH=0.1 -LOPCU=0 -LOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blomum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pcc -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USRR=US10089514 @CCN 1.1 6050 @runat_04102005_105744_8043 -NCPU=6 -ICPU=3
-DEV=MAP -LARGESUBSTRY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-NO_TIMEOUT=120 -MAIN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	585	16.4	698	5	BQ625203
2	576	16.1	894	7	CO024641
3	567	15.9	859	2	BE658136
4	566	15.8	768	2	CO980064
5	562	15.7	818	7	CO030342
6	561	15.7	820	7	CO034510
7	556.5	15.6	893	9	CG942457
8	553	15.5	529	6	CA003778
9	547.5	15.3	620	2	AW223881

10	541.5	15.2	645	7	CV099881	FAVU USDA
11	535.5	15.0	750	7	CO028609	CO028609 EST806993
12	525.5	14.7	579	7	CF598037	CF598037 NCST3e06
13	511	14.3	528	2	AW509018	AW509018 s13b01.y
14	500	14.0	890	7	CF703510	CF703510 CCAF0970
15	500	14.0	906	7	CF704316	CF704316 CCACN3070
16	486	13.6	1007	9	CNS06SL5	AL413391 T7 end of
17	483	13.5	523	8	B2412600	B2412600 OGAAG11TM
18	483	13.5	631	8	B2412593	B2412593 OGAAG11TC
19	483	13.5	631	8	B2537360	B2537360 OCGAD95TC
20	483	13.5	737	9	CG971361	CG971361 MBEK12TF
21	483	13.5	810	9	CG170785	CG170785 PUIIR58TD
22	473	13.2	658	7	CV128706	CV128706 Mdfrc3089
23	473	13.2	716	6	CD407015	CD407015 cm CK3199
24	472	13.2	540	4	BT427109	BT427109 s8F76e01.
25	468.5	13.1	773	7	CF718138	CF718138 CCAHBS0TF
26	450.5	12.6	1954	3	CNS0A87M	BX819469 Arabidops
27	450.5	12.6	2015	3	CNS0A7YD	BX820167 Arabidops
28	446.5	12.5	660	8	B2892725	B2892725 Hm7 0152.
29	443	12.4	1521	3	AY105009	AY105009 Zea mays
30	437.5	12.2	653	7	CF863831	CF863831 pZS008XG
31	436	12.2	845	8	B2570522	B2570522 m8r2_1411
32	433.5	12.1	502	4	BG508630	BG508630 sac75d07.
33	419	11.7	2450	3	AY104124	AY104124 Zea mays
34	417	11.7	657	9	CG947605	CG947605 MBEK88TF
35	415	11.6	643	2	BE823672	BE823672 pmt00021B
36	414	11.6	1259	8	B2557654	B2557654 pac81-60
37	399	11.2	1873	3	CNS0A1AN	BX829784 Arabidops
38	392.5	11.0	1043	8	B2548859	B2548859 pac81-60
39	389	10.9	655	9	CG919310	CG919310 MBEAL61TR
40	388.5	10.9	770	7	CF703416	CF703416 CCAAGD73TF
41	386	10.8	766	7	CF678009	CF678009 CCAHJ0TF
42	385	10.8	429	9	AG258434	AG258434 locus cor
43	384.5	10.8	668	6	CA095675	CA095675 SCCCL600
44	384.5	10.8	825	8	CC144065	CC144065 NDL. 75A11
45	383.5	10.7	679	7	CN926353	CN926353 000430AEP

ALIGNMENTS

RESULT 1
LOCUS BQ625203 698 bp mRNA linear EST 01-JUN-2002
DEFINITION USDA-FP_02294 Ridge pineapple sweet orange entire seedling Citrus
ACCESSION BQ625203
VERSION BQ625203.1 GI:21652372
KEYWORDS EST.
SOURCE Citrus sinensis
ORGANISM Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 698)

REFERENCE

Bausher, M., McKendree, W., Dang, P., Chaparro, J., Shatters, R.,
Hunter, W., and Niedz, R.
Expressed sequence tags isolated from entire sweet orange (C.
sinensis L. Osbeck) seedling
Unpublished (2003)

JOURNAL

Contact: Michael Bausher
US Horticultural Research
USDA - ARS
2001 South Rock Rd., Fort Pierce, FL 34945, USA
Tel: (772) 462-5918
Fax: (772) 462-5961
Email: mbausher@uhrl.ars.usda.gov
Seq primer: T3 Primer.

FEATURES

source location/Qualifiers
1..698
/organism="Citrus sinensis"
/mol_type="mRNA"
/cultivar="Ridge pineapple"
/db_xref="taxon:2711"

QY 564 HisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCys 583
 Db 373 CATCAGCTCGTCACACACATTCCACATCCGCCCAACGTCGGCGCTCCAGATT 314
 QY 584 ValArgAlaAlaPheProGlyGlySerMetThrGlyAlaProIlySlyArgThrMetGlu 603
 Db 313 CTCACAGCATGCTCTCCACACAGGATCATGACGGCGCTCCCAAACTCAGACCGCTGCAA 254
 QY 604 IleIleAspArgLeuGluGlu---GlyProArgGlyValIleSerGlyAlaLeuGlyTyr 622
 Db 253 ATCTTCACAGCGCTCGAAGAGACACCGAGCGCGCATATCTCCGACATCTGGGCTAT 194
 QY 623 PheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValIleAlaAsp 642
 Db 193 GTATGCCCGACGGCCACCGCTGATCGATCGCTGGTTATCCCTACGATTGTAACTACGGA 134
 QY 643 GlyGlnAlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGluGlu 662
 Db 133 AAGCAGCTTGAGCTGAGTGTGCTGAGGGCGCATTAATTGGCTCAGCGAGCGAGAAAGAA 74
 QY 663 PheThrGluThrValValIleValAlaArgAlaMetValThrAlaLeu 677
 Db 73 TGGGATGAGGTGATGTAAGCGAATGCTGTGCTACGGCGCTG 29
 RESULT 3
 BE658136/c
 LOCUS BE658136 859 bp mRNA linear EST 24-MAY-2001
 DEFINITION Gm70005A10A7 Gm-r1070 Glycine max cDNA clone Gm-r1070-1549 3',
 mRNA sequence.
 ACCESSION BE658136
 VERSION BE658136.1 GI:9984028
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max (soybean)
 Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 REFERENCE 1 (bases 1 to 859)
 AUTHORS Vodka, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Coryell, V.,
 Expanding, J., Rapp, C., Shoop, E., Pardinas, J., Liu, L. and Lewin, H.
 TITLE A Functional Genomics Program for Soybean (NSF 9872565)
 JOURNAL Unpublished (1999)
 COMMENT Other ESTs: A1940862 corresponding to Gm-cl010-1280 (5')
 Contact: Vodka, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l.vodka@uiuc.edu
 This clone is available through: Genome Systems, Inc. 4633 World
 Parkway Circle St. Louis, Missouri 63134. For further information
 call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
 427-3324 or contact: clone@genomesystems.com or info@genome
 systems.com web site: www.genomesystems.com
 Seq primer: 5'-TTTTTTTTTTTTTTT(A/C/G)-3'.
 FEATURES
 source
 1..859
 Location/Qualifiers
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1070-1549"
 /clone_lib="Gm-r1070"
 /note="The library Gm-r1070 is a sequence-driven, retracted
 set of 9,216 clones selected from cDNA libraries from
 various tissues and stages of development of soybean that
 represent 2,639 sequences from immature cotyledons, 1,770
 from immature seed coats, 3,938 from flowers, and 869
 from young pods. The 5' ESTs of the source clones from
 the different libraries was used to select singletons, or

a representative of each contig, which were retracted to
 form library Gm-r1070. The cDNA clones of the retracted
 Gm-r1070 library were then sequenced at the 3' end. The
 contig analysis to select unique genes was performed by
 the Laboratory of Ernest Retzel, Center for Computational
 Genomics and Bioinformatics, University of Minnesota,
 http://www.cbc.umn.edu/Research/Projects/Soybean/index.html
 .
 Retracting was performed by Genome Systems, St. Louis,
 http://www.genomesystems.com, and 3' sequencing by the
 Keck Center for Comparative and Functional Genomics,
 University of Illinois,
 http://www.life.uiuc.edu/dioech/keck.html. Note: The
 corresponding 5' EST from each clone in the Gm-r1070
 library is listed in the 'OTHER EST' field. The detailed
 information on the source library for each clone can also
 be obtained by referring to the Genome Systems clone ID of
 the original cDNA library that is also listed under
 'OTHER EST'."

ORIGIN
 Alignment Scores:
 Pred. No.: 4,42e-43 Length: 859
 Score: 567.00 Matches: 113
 Percent Similarity: 72.55% Conservative: 35
 Best Local Similarity: 55.39% Mismatches: 56
 Query Match: 15.87% Indels: 0
 DB: 2 Gaps: 0
 US-10-089-514-2 (1-686) x BE658136 (1-859)
 QY 474 GluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGly 493
 Db 797 GATCTGAGTANNNGCTNNNNNNNNNNAGAGGCTTCCTGATGATAGCAAAATATA 738
 QY 494 ValGluSerIlyProIleGlyGlyThrArgProArgGlyGlyValIleArgGluGlu 513
 Db 737 CTGAAGCTAAGCCCATCAAGGAGACGATGCTGCTGCTCTTAAGAGGAAGATGAG 678
 QY 514 ArgLeuArgAlaAspLeuAlaGlyArgGluIlyAspArgAlaGluLeuMetIleVal 533
 Db 677 CAACCTAAATTTAAATTACAGTTCACGCAAAAGATCAGGCTGAAACCTGAGATTGTT 618
 QY 534 AspLeuValArgAspLeuSerValValIleGlySerValIleValProArg 553
 Db 617 GACCTTCTAAGAGATACCTTGCGCGTGTGATCTGATCTGTCTATGTCACAGT 558
 QY 554 LeuPheGluValGluThrTyrAlaProValIleGlnLeuValSerThrIleArgGlyArg 573
 Db 557 CTCATGATGTGAATCATATGCACTTTCACACATAGTGATCTTCTGTTGGAAG 498
 QY 574 LeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlySerMet 593
 Db 497 AACGGTCAGATCTAGTGTGTAAGCTGTGTAAGAGCTGCAATTCCTGGTGTCAATG 438
 QY 594 ThrGlyAlaProIlySlyArgThrMetGluIleIleAspArgLeuGluGluProArg 613
 Db 437 ACAGGTCACANNATAGTTGAGATCAATGCAATCTTGATTCATTAAGATTCTTCGA 378
 QY 614 GluValIlySerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle 633
 Db 377 GGTATCTACACAGCGTATGATGATTTTCTCATTAATCAACACATTTGATTAATAT 318
 QY 634 ValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyAlaIle 653
 Db 317 GTGATAGAAGACGTATTGTACACAGAGGTAAGCTTCAATAGGAGCTGAGGCGCAAT 258
 QY 654 ValSerLeuSerAspGlnGluGluPheThrGluThrValValIlyAlaArgAlaMet 673
 Db 257 GTTGCTCTGTCAAAACCTGAAGAGATGAAGAGATGTTTGAAGACAAAGCCCA 198
 QY 674 ValThrAlaLeu 677
 Db 197 ACAAGGCTGTG 186

RESULT 4
 CO980064/c
 LOCUS
 DEFINITION
 CO980064 768 bp mRNA linear EST 13-SEP-2004
 Gm89007A1C03.r1 Gm-r1089 Glycine max cDNA clone Gm-r1089-2405 3,
 mRNA sequence.
 ACCESSION
 CO980064
 CO980064.1 GI:51336198
 VERSION
 EST.
 KEYWORDS
 Glycine max (soybean)
 SOURCE
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 768)
 REFERENCE
 Vodka, L., Shoemaker, R., Keim, P., Polacco, J. C., Retzel, E.,
 Khanna, A., Shealy, R., Clough, S., Thibaud-Nissen, F., Corryell, V.,
 Erpelidig, U., Gonzalez, D. O., Stromvik, M., Rodriguez-Huete, A. M.,
 Schweitzer, P., Gong, G. and Liu, L.
 A Functional Genomics Program for Soybean (NSF 9872565) (2004)
 Unpublished (2004)
 TITLE
 JOURNAL
 COMMENT
 Other_ESTRs: BE347129 corresponding to Gm-cl043-239 (5')
 Contact: Vodka, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodka@uiuc.edu
 Plate: Gm89007A1 row: C column: 03
 Seg primer: 5'-TTTTTTTTTTTTTTTTTTT(A/C/G)-3'
 High quality sequence stop: 768.
 Location/Qualifiers
 1..768
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1089-2405"
 /clone_1b="Gm-r1089"
 /note="The library Gm-r1089 is a sequence-driven, reracked
 set of 9,216 low redundancy clones selected from 38
 different cDNA libraries constructed from various tissues
 and stages of development of soybean including 973 cDNAs
 from germinating cotyledons (source library Gm-cl069,
 Gm-cl076, and Gm-cl077); 1,465 cDNAs from various tissue
 and organ systems of the adult plant; 476 cDNAs from adult
 stem tissue (source library Gm-cl062); 1340 cDNAs from
 tissue culture derived somatic embryos (source libraries
 Gm-cl036 and Gm-cl075); 2918 cDNAs from hypocotyls or
 young seedlings; 742 cDNAs from germinating seedlings,
 shoot tips, or leaves exposed to various stresses (source
 libraries Gm-cl065, Gm-cl066, and Gm-cl068); 839 cDNAs
 from young leaves or hypocotyls exposed to bacterial and
 fungal pathogens (source libraries Gm-cl072, Gm-cl073,
 Gm-cl074; and Gm-cl084); and 463 from roots of young
 plants grown in hydroponic media without phosphate (source
 library Gm-cl087). The 5' ESTs of the source clones from
 the different libraries were used to select singletons, or
 a representative of each contig, which were reracked to
 form library Gm-r1089 and the cDNA clones of the reracked
 Gm-r1089 library were then sequenced at the 3' end. The
 unigene selection and 3' sequencing was funded by NSF
 Plant Genome project #9872565
 (http://soybeanomics.croscs.uiuc.edu/) as part of
 creation of a low redundancy soybean cDNA set. The source
 cDNA libraries were constructed by the Laboratory of
 Lila Vodka, University of Illinois, Randy Shoemaker, Iowa
 State University, and Paul Keim, Northern Arizona
 University as part of the Public EST project,
 http://129.186.26.94/soybeanest.html. The contig analysis

to select unique genes was performed by the laboratory of
 Ernest Retzel, Center for Computational Genomics and
 Bioinformatics, University of Minnesota,
 http://web.abc.umn.edu/biodata/soy//. Reracking and 3'
 sequencing were conducted by services of the University of
 Illinois Keck Center for Comparative and Functional
 Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The
 corresponding 5' EST from each clone in the Gm-r1089
 library is listed in the 'OTHER EST' field. The detailed
 information on the source library for each clone can also
 be obtained by referring to the clone ID of the original
 cDNA library that is also listed under 'OTHER EST'."

ORIGIN

Alignment Scores:

Prod. No.:	4,66e-43	Length:	768
Score:	566.00	Matches:	113
Percent Similarity:	73.04%	Conservative:	36
Best Local Similarity:	55.39%	Mismatches:	55
Query Match:	15.84%	Indels:	0
DB:	7	Gaps:	0

US-10-089-514-2 (1-686) x CO980064 (1-768)

Qy	474	GlueUSeRValleuSeRAlaSeRProGluArgPheUThrIleGlyAlaAspGlyGly	493
Db	745	GATCTGATATTGTCGTTCTTNNNTGAGAGGTCCTCAGNNNGATGGAATATA	686
Qy	494	ValGluSeRlyeProIleYsglyThraGProaRgGlyGlyThraIaGluGluAspGlu	513
Db	685	CTAGAAGCTTAAGCCCTCAAGGAAGATAGCTCGTGCTCACTGAAGAGAAGATGAG	626
Qy	514	ArgLeuValAlaAspLeuAlaGlyArgGluYAspArgAlaGluAsnLeuMetIleVal	533
Db	625	CACCTAAATTAAATTAACTGACAGGAAAGATCAGGCTGAAAGCTGATGTTGTT	566
Qy	534	AspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArg	553
Db	565	GACCTTCTAAGGATGATCCTTGCGCGTGTGTTGATCCTGATGCTGATGCGACGT	506
Qy	554	LeuPheGluValAlaGluThryAlaProValHisGluLeuValSerThryIleArgGlyArg	573
Db	505	CTCATGATGATGGAATCATATGCACTGTTCAACATGATGATGATGATGATGATG	446
Qy	574	LeuArgProGlyThySerThrAlaIaCysValArgAlaAlaPheProGlyGlySerMet	593
Db	445	AAGCGTCAAGATGCTGCTGATGATGATGATGATGATGATGATGATGATGATG	386
Qy	594	ThrGlyAlaProLysLysArgThyMetGluIleIleAspArgLeuGluGluProArg	613
Db	385	ACAGTGACACCTTAAGTGAATGATGATGATGATGATGATGATGATGATGATG	326
Qy	614	GlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle	633
Db	325	GGATCTACTCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATG	266
Qy	634	ValIleArgThrIleValLeuAlaAspGlyGluAlaGluPheGlyValGlyGlyAlaIle	653
Db	265	GGATTAAGACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG	206
Qy	654	ValSerLeuSerAspGluGluGluGluPheThrGluThryValValIleAlaArgAlaMet	673
Db	205	GTTTGCTCTGCAACCTCAAGACGATGATGATGATGATGATGATGATGATGATG	146
Qy	674	ValThrAlaLeu 677	
Db	145	ACAAGGCTGTG 134	

RESULT 5

CO030342/c
 LOCUS
 DEFINITION
 CO030342 818 bp mRNA linear EST 10-JUN-2004
 EST080726 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
 kb Coccidioides posadasii cDNA clone C1F368 3' end, mRNA sequence.

Query Match:	15.70%	Indels:	4
DB:	7	Gaps:	3
US-10-089-514-2 (1-686) × C0034510 (1-820)			

429 IleArgSerGluGlySerProGluIleCysLeuThrAspMetValThrAlaPro---Thr 447
 787 ATTAAGGAGGGCGAGCGCATGTAGGTGAGCAATCACCAAAATTCAAGCATGTGTCCG 728
 448 GluAlaThrAlaLeuProLeuTyrSerAlaLeuArgAlaIleSerProValProTyrGly 467
 727 GATGTAGAACCCATAGCATTTATATCTCTCCCTTCGAGAAAGTAAACCCGCCCTTACTCA 668
 468 AlaLeuLeuGluPhe-----ProGluLeuSerValLeuSerAlaSerProGluArgPhe 485
 667 GCGTATATAGACTCTCCGTGTCAACGAGACCATCATCTCTTCGTGTGCGCCGAGAGATT 608
 486 LeuThrIleGlyAlaAspGlyGlyValIleGlySerProIleGlyGlyThrArgProArg 505
 607 ATATCATATGATGTCTGATGTCGCTGCTGCTGTAATGAAACCAATTAAAGGAACCTTGGCCTC 548
 506 GlyIlyThrAlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluAsp 525
 547 AGCCAGACAAGAAAGAAACGAGCGGAGAAAGCAAGTACGATGACGAGTGCAGAGAG 488
 526 ArgAlaGluAspLeuMetIleValAspLeuValArgAspAspLeuAspSerValCysAla 545
 487 CTGCGCGGAGAACTTATATGATTGTTCGATCTTATTCGCGCGGACCTCCACAAATTTCTCCA 428
 546 IleGlySerValHisValProArgLeuPheGluValGluThrThrAlaProValHisGln 555
 427 TCAAAATCATATAGAGTCCCGAAACTACTCCACGTTGAAAGCTACCAAAACAGTACATCAG 368
 566 LeuValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArg 585
 367 CTCGTCAACAACATTCATCAATCCACATCGGCCCAAGTGGCGGCGTCCAGCTTCTCGAG 308
 586 AlaAlaPheProGlyGlySerMetThrGlyAlaProGlyLysArgThrMetGluIle 605
 307 CGATGCTTCCCAACAGATCATGTACGGGCGCTCCCAAACTCAGAGCCCGCAAAATCTTC 248
 606 AspArgLeuGluGlu---GlyProArgGlyValTyrSerGlyValaLeuGlyTyrPheAla 624
 247 GACGGCCTGTGAAGAGACCGCAGCGCGGATATATCTCCGGCATGTGGGCTTATGTATGC 188
 625 LeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGln 644
 187 GCCACGGGACCGTGTGATCAGTGTGCGTTATCTCCGATCTTAAAGTAAAGTACGAAAGCAG 128
 645 AlaGluPheGlyValGlyGlyAlaIleValSerLeuSerAspGlnGluGluPheThr 664
 127 CTTGAGCTAGAGTGTGAGAGGGCGATTAATTACTGTGCTCAGCGAGCGAGAAAGAAATGGAT 68
 665 GluThrValValLeuAlaArgAlaMetValThrAlaLeu 677
 67 GAGGTCAATGTGAAGCGCAATGCTGTGTCTACGCGCTG 29

LOCUS	DEFINITION	CG942457	893 bp	DNA	linear	GSS 15-DEC-2003
CG942457	MBE0909TF mch2 Medicago truncatula genomic clone 57P11, survey sequence.					

ACCESSION	CG942457
VERSION	CG942457.1
KEYWORDS	GI:39847070
SOURCE	GSS.
ORGANISM	Medicago truncatula (barrel medic)
	Medicago truncatula

REFERENCE
TOWN, C. D., SHETY, J., KOO, H. and FELDLIYUM, T. F.
1 (bases 1 to 893)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eustosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 893)

TITLE	Sequencing of BAC ends from Medicago truncatula
JOURNAL	Unpublished (2003)
COMMENT	Other_GSSs: MBEHO90TR

9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-3523

Email: cdtown@tigr.org
Seq primer: TGTAAACGACGGCCAGT
Class: BAC ends.

FEATURES	Location/Qualifiers
source	1. .893

```

/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone="57p11"
/clone_1b="mth2"
/notes="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Cook, D.R. and Kim, D.J., unpublished"

```

Alignment Scores:		
pred. No.:	4,656-42	length:
Score:	556.50	Matches:
Percent Similarity:	56.00%	Conservative:
Best Local Similarity:	43.33%	Mismatches:
Query Match:	15.56%	Indels:
DB:	9	Gaps:
US-10-089-514-2 (1-686)	x CG942457 (1-893)	

US-10-089-514-2 (1-686) X CG942457 (1-893)

Qy	378	ThleuThhGlyleuAlaValAlaArgAlaProAlaGluProThrProAlaMetValPheGly	397
Db	9	AGCTTAACGTGG---TCGTGAGAGATGGATTATGAAAGCAGCACTCTCATCTTCAACT	65
Qy	398	lleProGluAlaAlaAlaGlyPheGlyProleuAlaArgAlaArgHisAspLeuAspAla	417
Db	66	TTTTCCCAATAAAGCTGGTTTTGCCA-----GCTGAATAATCTAAAGACAG	113
Qy	418	TyrLeuLysArgGlyLeuAspGluCysLeuLysGluIleArgAsnGlyGlySerTyrGluIle	433
Db	114	TACATTAGAGATGTTTAAAGATGTCCTAAACATACATTAGATGAGAGAGAGACTATGAGTTG	172
Qy	438	CysLeuThrAsnMetValThrAlaProThrGluAla---ThrAlaLeuProLeuTyrSer	456
Db	174	TGCTCTACACCAACCGATTAAGAAACCGGTGAGTATTAAATCTCTTGGACTTTACCTA	233
Qy	457	AlaLeuArgAlaAlaSerProValProTyrGlyAlaLeuLeuGluPhePro-----Glu	474
Db	234	CATTGAGGAAAGAAATCCACACACCTTATGGGGCTGGCTGAATTTTCCAAAGAAAGAT	293
Qy	475	LeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAspGlyGlyVal	494
Db	294	CTGTGATTTGCTGTTCTTCCCCGAGAGGTTCTCGACGTTGGATGAGAGTATGCTA	353
Qy	495	GluSerLysProIleLysGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArg	514
Db	354	GAACTMAACCCATTAAAGGTACTATTAAGCTCGTGGTGTACTGGAAGAGAAACAAACA	413
Qy	515	LeuArgAlaAspLeuAlaGlyArg-----	522
Db	414	CTCAAAATTGAATTTACAGCTCAGGTTTTTCTATGACTTCAATTATCATATTGGAAATTGCT	473
Qy	522	-----	522
Db	474	TAACTTACTACATTTTTTTGTATGACATCAAAATTTCCAAGCACTGATTTGAATCTGTG	533
Qy	523	-----GluLysAspArgAlaGluAsnLeuMetIleVal	533
Db	534	CGAAGCTGTCGTTTTTTTCTTTCAGTAAAGATCCAGGACGAAACCTGATATATGTTT	593

QY 534 AspleuValArgAnaAspleuAsnSerValCysAlaIleGlySerValHisValProArg 553
Db 594 GACCTTCAAGAAATGACCTTGCTGCTATGATCCTGGATCTGTTGACGTACCCGAT 653
QY 554 LeupheValValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArg 573
Db 654 CTATGGAGAGACATATATGACCACTGTCACCAACATGGAGATGACATTCGTGGGAAA 713
QY 574 LeuArgProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMet 593
Db 714 AAGAGGTCAAGATATAGTCTGTGACTGTCTCAAGGTGCATTTCCGGGTGTCATG 773
QY 594 ThrGlyAlaProGlyValArgThrMetGluIleIleAspArgLeuGluGlyProArg 613
Db 774 ACAGGCCACCAAGTGAAGTCAATGGAACCTTCTGCTCGAAAGTGTTCTCGA 833
QY 614 GlyValIleSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIle 633
Db 834 GGCATCTACTACAGCTCATTGGATT-TTCTCATATATATCAACATTTGATCTAATATTT 892

RESULT 8
CA003778 529 bp mRNA linear EST 23-OCT-2002
LOCUS HS15114r HS Hordeum vulgare subsp. vulgare cDNA clone HS15114
DEFINITION 5-PRIME, mRNA sequence.
ACCESSION CA003778
VERSION CA003778.1 GI:24280760
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare
ORGANISM Hordeum vulgare subsp. vulgare
REFERENCE Bukaryova; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Poideae; Triticeae; Hordeum.
1 (bases 1 to 529)
Zhang,H., Potokina,E., Michalek,W., Weesche,W., Stein,N. and
Graner,A.
Barley ESTs from germinating seeds
Unpublished (2002)
CONTACT: Stein Nils
Molecular Markers Group, Department Genbank
Institute of Plant Genetics and Crop Plant Research (IPK)
Corrensstr. 3, 06466, Gatersleben, Germany
Tel: 039482-5522
Fax: 039482-5595
Email: stein@ipk-gatersleben.de
Insert Length: 529 Std Error: 0.00
Plate: 15 row: 1 column: 14
Seq primer: M13rev.
FEATURES
source location/Qualifiers

1..529
/organism="Hordeum vulgare subsp. vulgare"
/mol_type="mRNA"
/cultivar="barke"
/sub_species="vulgare"
/db_xref="GABI:258990"
/db_xref="taxon:112509"
/clone="HS15114"
/tissue_type="embryo + scutellum"
/dev_stage="0-16 hours after imbibition"
/lab_host="XL10-Gold"
/clone_lib="HS"
/note="Vector: pBluescript SK+; Site 1: EcoRI (5'-end of cDNA); Site 2: XhoI (3'-end of cDNA); Due to a cloning artefact caused by the kit, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter used for cloning. To excise the insert, restriction sites upstream EcoRI should be used (e.g. BamHI, SalI, PstI). NOTE: Also due to the cloning system used Blue/white selection for recombinants is not 100% reliable."

ORIGIN
Alignment Scores:
Pred. No.:

4.61e-42 Length: 529

Score: 553.00 Matches: 109
Percent Similarity: 76.00% Conservative: 24
Best Local Similarity: 76.00% Mismatches: 42
Query Match: 15.48% Indels: 0
DB: 6 Gaps: 0
US-10-089-514-2 (1-686) x CA003778 (1-529)

QY 496 SerIleProIleIleGlyThrArgProArgGlyGlyThrAlaGluGluAspGluArgLeu 515
Db 5 ACAGGCCAATCAAGGTATCAATATGACACCTGGAGAACACCAAGAGATGATGTCCTA 64
QY 516 ArgAlaAspLeuAlaGlyArgGlyValAspArgAlaGluAsnLeuMetIleValAspLeu 535
Db 65 CGTTTGCACTGAAATACAGTAAAGACACCAAGCTGAGAACTTGATGATTTGATCTC 124
QY 536 ValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisValProArgLeuPhe 555
Db 125 TTAAGAAATGATCTGGGTAGTCTCGACCTGGAGCGGTGATGTTCTCGCTCATG 184
QY 556 GluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArgGlyArgLeuArg 575
Db 185 GAGGTGAAATCAATATATCTGTTCAACCATGTGTGACCATCCGCGAAGAAAGAAC 244
QY 576 ProGlyThrSerThrAlaAlaCysValArgAlaAlaPheProGlyGlySerMetThrGly 595
Db 245 CTGATCTTAAGCCCGGTAGACTGCATCAAGCCGCTTTCAGAGAGCTGATGACGGGT 304
QY 596 AlaProIleValArgThrMetGluIleIleAspArgLeuGluGlyProArgGlyVal 615
Db 305 GCCCGAAGGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 364
QY 616 TyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIle 635
Db 365 TACTCCGGGTCGATCCGGGTCTTCTGTCACCAACCGCATTCGATGAAATCGATCATC 424
QY 636 ArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGlyValAlaIleValSer 655
Db 425 AGACGGTATGCTGACGACGCGGTGCTCGCTCGCTCGCTCGCTCGCTCGCTCGCTCG 484
QY 656 LeuSerAspGlnGluGluPheThrGluThrValIleValAla 670
Db 485 CTGTCAAGCCCGAGCGGAGTACGCCGAGATGATCTCAAGCG 529

RESULT 9
AM223881 620 bp mRNA linear EST 18-MAY-2001
LOCUS EST100692 tomato fruit red ripe, TAMU Lycopersicon esculentum cDNA
DEFINITION clone cLENI3P22, mRNA sequence.
ACCESSION AM223881
VERSION AM223881.1 GI:6535565
KEYWORDS EST.
SOURCE Lycopersicon esculentum (tomato)
ORGANISM Lycopersicon esculentum
REFERENCE Bukaryova; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; lamids; Solanales; Solanales; Solanales; Solanum; Lycopersicon.
1 (bases 1 to 620)
Alcala,J., Vrbalov,J., White,R., Matern,A.L., Holt,I.E., Liang,F.,
Upton,J., Hansen,T., Craven,M.B., Bowman,C.L., Ahn,S.,
Romling,C.M., Frazer,C.M., Martin,G.B., Tanksey,S.D. and
Giovannoni,J.
Generation of ESTs from tomato fruit tissue
Unpublished (1999)
CONTACT: CUGI
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Email: <http://www.genome.clemson.edu/orders/index.html>
5 prime sequence.
FEATURES
source location/Qualifiers
1..620
/organism="Lycopersicon esculentum"

```

/mol_type="mRNA"
/cultivar="TA496"
/db_xref="taxon:4081"
/clone="cLEN13P22"
/tissue_type="pericarp"
/dev_stage="red ripe (7-20 days post-breaker)"
/clone_id="tomato fruit red ripe, TAMU"
/notes="Vector: pBluescript SK(-); Site 1: EcoRI; Site 2:
XhoI; supplier: Giovannoni; Fruit were tagged at the
breaker stage (first sign of lycopene accumulation on the
blossom end of the fruit) and harvested 7 days
post-breaker (fully red-ripe), 10 days post breaker, and
20 days post-breaker (over-ripe). 20 day fruit which
showed external or internal signs of pathogenesis were
discarded. Fruit were cut in half and the seeds and
locules were discarded prior to freezing the pericarp."

```

ORIGIN

Alignment Scores:

Pred. No.:	1,94e-41	Length:	620
Score:	547.50	Matches:	112
Percent Similarity:	67.96%	Conservative:	28
Best Local Similarity:	54.37%	Mismatches:	63
Query Match:	15.32%	Indels:	3
DB:	2	Gaps:	2

US-10-089-514-2 (1-686) x AM223881 (1-620)

```

QY 419 LeuLYeArGILeSpJuCySLeuLySGuILleArGAsngLYGUsSeTyrgIuILeCyS 438
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 3 ATCAAAAGACGTTGATGATTCATGAAAGATTCAAAAGAAAGATTATGAGTGTGT 62
   ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 439 LeuThrAsmeValThrAlaProThrGLuAla---ThrAlaLeuProLeuTySerAla 457
   ||||| ::::: ::::: ::::: ::::: ::::: ::::: :::::
DB 63 CTACAACTCAGATGAGAAATGAGATTGGGGGGAATAGATTCTCGAATCTTATCGTAA 122
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 458 LeuArGAlaIleSerProValProTyrgLYAlaLeuLeuGluDhe-----ProGluLeu 475
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
DB 123 CTCGAAATAGAAATATCTCGCATATGCTGCTGCTTATTTTCAAGGAGAAACCTTA 182
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 476 SerValLeuSerIaSerProGluArgPheLeuThrIleGlyAlaSepGlyValGlu 495
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
DB 183 AGCATATGTTGTCATCACCCTGAAGGTTCTTACGATTCGACGAGATGCTATTTAGAA 242
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 496 SerLYeProIleLySGlyThrArgProArgLYGlyThrAlaGluIaSepGluArgLeu 515
   ::::: ::::: ::::: ::::: ::::: :::::
DB 243 GCAAAACCCATAAAGGACTATAGCTCGGTCCACCCCAAGAGAAATGAAATTTCTG 302
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 516 ArgAlaSepIeuaIaGlyArgGluLyAspArgAlaGluAsnLeuMetIleValaSepLeu 535
   ::::: ::::: ::::: ::::: ::::: :::::
DB 303 AAACCTCAATTAGAAATGACGATAAAGATCAGCGGAAATTTGATGATTGTTGACTTG 362
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 536 ValArgAsnAspIeuaSerValIleGlySerValHisValProArgLeuPhe 555
   ::::: ::::: ::::: ::::: ::::: :::::
DB 363 TTGAGAAATGACCTTGCGGTGATGTGAGACTGCGCTGTTGATGTCCTCAATCTCATG 422
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 556 GluValaGluThrTyraIaProValHisGluLeuValSerThrIleArgGlyArgLeuArg 575
   ::::: ::::: ::::: ::::: ::::: :::::
DB 423 GAAATGAAATCCATGTCACAGATTTCATCAGTGTGATTCGCGGGGAAAGCGCA 482
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 576 ProGlyThrSerThrAlaAlaCysValaArgAlaAlaPheProGlyGlySerMetThrGly 595
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
DB 483 TCAGATGCAAGTCAATTGATGTGTGTTAGTGCAATTCCTGCTGTGTGATGACAGG 542
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 596 AlaProLyLeuArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyVal 615
   ::::: ::::: ::::: ::::: ::::: :::::
DB 543 GCACCAAGTTGATGCAATGAACTTTGATCATCTTGAAATTTGTTGAGGGGCGCAT 602
   ||||| ::::: ::::: ::::: ::::: ::::: :::::
QY 616 TyrSerGlyAlaLeuGly 621
   ||||| ::::: ::::: ::::: ::::: :::::
DB 603 TACTCGGGCTGCTATTTGA 620
   ||||| ::::: ::::: ::::: :::::

```

RESULT 10

CV099881
LOCUS
DEFINITION
FAMU USDA FP 7904 Vitis shuttleworthii L., grape Vitis
shuttleworthii cDNA clone WNVs046_D11 5', mRNA sequence.
CV099881
ACCESSION
VERSION
CV099881.1 GI:51582045
KEYWORDS
SOURCE
ORGANISM
Vitis shuttleworthii
Vitis shuttleworthii
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosids; Vitaceae; Vitis.

REFERENCE

AUTHORS
Hunter, W.B., Dang, P.M., Chaparro, D.X., Lu, J., Huang, H. and Leong, S.
Genes expressed in Vitis shuttleworthii L
Unpublished (2004)
COMMENT
Contact: JIANG Lu, Wayne Hunter, Phat Dang and Hong Huang
Genetic Lab, Viticulture Center; U.S. Horticultural Research Lab
FAMU, USDA-ARS;
Tallahassee, FL; Ft. Pierce, FL, USA
Tel: (850) 412-7393, (772) 462-5898
Fax: (850) 561-2617, (772) 462-5986
Email: jiang.lu@fam.u.edu, hong.huang@fam.u.edu,
whunter@usnrl.ars.usda.gov, pdang@usnrl.ars.usda.gov
Seq primer: T3 Primer.

FEATURES

source

```

1..645
Location/Qualifiers
1..645
/organism="Vitis shuttleworthii"
/mol_type="mRNA"
/db_xref="taxon:246827"
/clone="WNVs046_D11"
/sex="Mixed population"
/tissue_type="Entire tendril, leaves, bud, flowers"
/dev_stage="At blooming"
/lab_host="Xli-Blue"
/clone_id="Vitis shuttleworthii L., grape"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; A high quality EST with at least 100 contiguous
bases at Trace Turner score of 20 or better. Construction
by PM Dang, USDA, ARS, U.S. Horticultural Research Lab,
Ft. Pierce, FL, USA."

```

ORIGIN

Alignment Scores:

Pred. No.:	7,65e-41	Length:	645
Score:	541.50	Matches:	117
Percent Similarity:	70.56%	Conservative:	34
Best Local Similarity:	54.67%	Mismatches:	60
Query Match:	15.16%	Indels:	4
DB:	7	Gaps:	2

US-10-089-514-2 (1-686) x CV099881 (1-645)

```

QY 434 SerTyrgIuILeCySeuThrAsmeValThrAlaProThr--GluAlaThrAlaLeu 452
   ||||| ::::: ::::: ::::: ::::: :::::
DB 3 AGTTAAGAATGTGCTTCAACTCAATGAGAAAGAGATTGGCAATGATTACTTG 62
   ||||| ::::: ::::: ::::: ::::: :::::
QY 453 ProLeuTySerAlaLeuArgAlaIleSerProValProTyrgLYAlaLeuLeuGluPhe 472
   ||||| ::::: ::::: ::::: ::::: :::::
DB 63 GGACTTATCTTAACCTTGAAGAAAAAATCCAGCACATATGACACCTGCTTAATTTT 122
   ||||| ::::: ::::: ::::: ::::: :::::
QY 473 ProGlu-----LeuSerValLeuSerIaSerProGluArgPheLeuThrIleGlyAla 490
   ::::: ::::: ::::: ::::: :::::
DB 123 TCAGAACAAAACCTGTCGATCTGCTGCTTTCACAGAGAGTTTTCACGCTGATGCA 182
   ||||| ::::: ::::: ::::: ::::: :::::
QY 491 AspGlyGlyValGluSerLyAspProIleLySGlyThrArgProArgGlyThrAlaGlu 510
   ::::: ::::: ::::: ::::: :::::
DB 183 AATGTATTTTAGAGCAAAAGCCCATCAAGGGTACTATGTCGTGTTGAGAAAGAG 242
   ||||| ::::: ::::: ::::: ::::: :::::
QY 511 GluAspGluArgLeuArgAlaSepIeuaIaGlyArgGluLyAspArgAlaGluAsnLeu 530
   ||||| ::::: ::::: ::::: ::::: :::::
DB 243 GAAGATGAACACCTCAAAATTTGCAATCAATACAGTGAAGAAAGATCAGGCTGAATCTG 302
   ||||| ::::: ::::: ::::: :::::

```


QY 531 MetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHis 550
 Db ATGATGCTGATCTTCTTAAGAAATGACCTTGCTGCTGTAACCTGGAATCTATTCAT 362
 QY 551 ValProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIle 570
 Db GTTCTCTGCTTATGATGGAATGGAATCATCACTGTTCAATCAACGATGATGATCAT 422
 QY 571 ArgGlyLeuArgProGlyThrSerThrAlaIleCysValArgAlaIleAlaPheProGly 590
 Db CGAGAAAGAAAGCAATCAAGATGAGCCCACTGATGTGTCAGAGAGACCTTCCAGGT 482
 QY 591 GlySerMetThrGlyAlaProValArgThrMetGluIleIleAspArgLeuGluGlu 610
 Db GGATCATGACAGGTGCGCCGCAAGTTAAGATCATGAACTCTGATTCATTGAAAT- 541
 QY 611 GlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAsp 630
 Db AGTTCTGAGGTATCTACTCTGCTGCTGAGGTTTTCATCAACACAGACATTGGAT 601
 QY 631 LeuSerIleValIleArgThrIleValLeuAlaAspGlyGln 644
 Db CTCAACATGTGTATGAGACGATTTGATCATGAGGTGAA 643

RESULT 11
 LOCUS CO028609/c 750 bp mRNA linear EST 10-JUN-2004
 DEFINITION EST806999 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
 ACCESSION CO028609 Coccidioides posadasii cDNA clone ClFAS82 3' end, mRNA sequence.
 VERSION CO028609.1 GI:48559837
 KEYWORDS EST.
 SOURCE Coccidioides posadasii
 ORGANISM Coccidioides posadasii
 Bacteria; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Onygenales; mitosporic Onygenales; Coccidioides.
 REFERENCE 1 (bases 1 to 750)
 AUTHORS Gardner M.J. and Cole G.T.
 TITLE Analysis of gene expression in Coccidioides posadasii mycelia and
 spherules via expressed sequence tags
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: EST806994
 Contact: Gardner MJ
 The Institute for Genomic Research
 9112 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: gardner@igf.org

FEATURES
 source location/Qualifiers
 1..750
 /organism="Coccidioides posadasii"
 /mol_type="mRNA"
 /strain="C735"
 /db_xref="taxon:199306"
 /clone="ClFAS82"
 /dev_stage="spherules"
 /lab_host="E. coli DH10B, T1 phage resistant"
 /clone_lib="Coccidioides posadasii spherule cDNA library,
 0.5 to 5.3 kb"
 /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV;
 Coccidioides posadasii spherule cDNA library, 0.5 to 5.3
 kb"

ORIGIN

Alignment Scores:
 Pred. No.: 3,55e-40 Length: 750
 Score: 535.50 Matches: 115
 Percent Similarity: 66.08% Conservative: 35
 Best Local Similarity: 50.66% Mismatches: 74
 Query Match: 14.99% Indels: 3
 DB: 7 Gaps: 2

US-10-089-514-2 (1-686) x CO028609 (1-750)

QY 454 LeuTyrSerAlaLeuArgAlaIleSerProValProTyrGlyAlaLeuLeuGluPhe--- 472
 Db TTATATCTTCCTTCAGAAAGTAACTGACCCCTTACTACAGGATATATATCTTCCT 650
 QY 473 ---ProGluLeuSerValLeuSerAlaSerProGluArgPheLeuThrIleGlyAlaAsp 491
 Db GTCAACAGAACCAACATACCTCTTCCTGCTCCGAGAGGTTTATATCATAGATGCTGAT 590
 QY 492 GlyGlyValAlaLeuSerLysProIleGlyTyrArgProArgGlyGlyThrAlaGluGlu 511
 Db GGCGTGCTGAATAATGAACCAATATAAGGAACCTTGCCCTGACCCAGACCAAGAAAGAA 530
 QY 512 AspGluArgLeuArgAlaAspLeuAlaGlyArgGlyLysAspArgAlaGluLeuMet 531
 Db GACGAGCGAGAGAAAGTCAAGTACGACGATGTCAAGAGAGCTCGCGGAACTTATATG 470
 QY 532 IleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIleGlySerValHisVal 551
 Db ATGTGATCTTATTCGCGCGGACCTCCACAACTTCTTCATCAAAATCAATCAAGTTC 410
 QY 552 ProArgLeuPheGluValGluThrTyrAlaProValHisGlnLeuValSerThrIleArg 571
 Db CCGAACTACTCCACGCTGAAAGCTACGAAACAGTACATCAGCTCGTCACACACATTCAA 350
 QY 572 GlyArgLeuArgProGlyThrSerThrAlaIleCysValArgAlaAlaPheProGlyGly 591
 Db TCCCAATCCGCCCCCAACGTCGCGCGCTGAGTTCACAGATGCTTCCACACAGGA 290
 QY 592 SerMetThrGlyAlaProValArgThrMetGluIleIleAspArgLeuGluGlu--- 610
 Db TCCATACGCGCGCTCCCAACTCAGAGCCGTGCAATCTCCACGCGCTGGAAGACAC 230
 QY 611 GlyProArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeuSerGlyAlaAlaAsp 630
 Db CGGACGCGCGCATATATCTCGGCACTGAGCTATGTATGCCAGCGGACCGTCGAT 170
 QY 631 LeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAlaGluPheGlyValGly 650
 Db CACTCGGTGTTATCCGTACGATTTGAAATGACGGAACAGCTTATGATGCTGGA 110
 QY 651 GlyAlaIleValSerLeuSerAspGlnGluGluPheThrGluThrValValAla 670
 Db GGAGCCATTTACTTGCTACGCGGAGGAGAAAGATGAGTACTGATGTAAGGCG 50
 QY 671 ArgAlaMetValThrAlaLeu 677
 Db AATGCTGTGTACGCGCGCTG 29

RESULT 12
 LOCUS CF598037 579 bp mRNA linear EST 26-SEP-2003
 DEFINITION NCES73e06c12.v1 Ne-Liv Tachyzoite cDNA library Neospora caninum
 cDNA 5' similar to TR:P72539 P72539 PAPA.;, mRNA sequence.

ACCESSION CF598037 GI:36356095
 VERSION EST.
 KEYWORDS Neospora caninum
 SOURCE Neospora caninum
 ORGANISM Neospora caninum

REFERENCE 1 (bases 1 to 579)
 AUTHORS Cole R., Fogarty S., Tang K., Howe D.K., Sibley L.D., Clifton S.,
 Maira M., Hillier L., Pape D., Martin J., Wylie T., Theising B.,
 Bowers Y., Gibbons M., Ritter E., Bennet J., Ronko I.,
 Teagarden J., R., Fedele M., Belaygorod L., Franklin C.,
 Carr L.M., Grow A., Maguire L., Wadkins J., Richey J., Waterson R.
 and Wilson R.
 USDA-Washu Neospora EST Project
 Unpublished (2000)
 Contact: Sandy Clifton, Ph.D. - Neospora
 USDA-Washu Neospora EST Project
 Washington University School of Medicine

TITLE
 JOURNAL
 COMMENT

Washington University School of Medicine

Oy 499 llelvsglyThrArgProarglyglYThrIaagiuglaaBgluaBgluaBgluaBp 518
 Db 1 ATCAAGGAAACGATGCTCTGTGGTCTACTGAAGAGAAAGATGACCACTAAAAATTAAA 60
 Oy 519 LeuAlaGIArgGluLysAspArgIaGluAsnLeuMetIleValAspLeuValArgAsn 538
 Db 61 TTACAGTTCAGGAAAGATGACCTGAAACCTGATGATTGTGACCTTTAGGAAT 120
 Oy 539 AspleuAsnSerValCysAlaIleGlySerValHisValProArgLeuPheGluValGlu 558
 Db 121 GACCTTGCGCCGGTGTGTGATCCTGATCCTGATCTGTTCAATGTCACGCTCATGATGTGGA 180
 Oy 559 ThrTyraIaProValHisGlnLeuValSerThrIleArgGlyArgLeuArgProGlyThr 578
 Db 181 TCATATGCAACTGTTCAACAATGATGTGATCTATTCTGGGCAAAAGCGGTCAAGATGTC 240
 Oy 579 SerThrAlaIaCysValArgValaIaIaPheProGlyGlySerMetThrGlyValaBpLys 598
 Db 241 AGTGTGTAGACTGTGTCAAGCTGATTTCTGGTGTTCATATACAGGTGACCTTAAG 300
 Oy 599 LysArgThrMetGluIleIleAspArgLeuGluGluGlyProArgGlyValaTySerGly 618
 Db 301 TTGAGATGCAAGGAACCTTGATCTCATGTGAATGTCTCCAGAGATCTACCAAGC 360
 Oy 619 AlaIeuGlyTTPheAlaLeuSerGlyAlaAlaAspLeuSerIleValIleArgThrIle 638
 Db 361 TGATGTGATTTTCTCATATATATCAACGATTTGATCTGAATATGTGATTAAGACAGTC 420
 Oy 639 ValIeuAlaAspGlyGlnAlaGluPheGlyValGlyGlyAlaIleValaSerLeuSerAsp 658
 Db 421 ATTGTACACGAGGGGGAAGCTTCATATAGAGCTGACGGCAGCATGGTGTCTGTCAAC 480
 Oy 659 GlnGluGluGluPheThrGluThrValValLysAlaArgAla 672
 Db 481 ACTTGACGACGATATGAAGAGTGGGATGGAAGACACACAGCG 522

ORGANISM	Cryptococcus neoformans var. neoformans Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella.
REFERENCE	1 (bases 1 to 890)
AUTHORS	Loftus, B.
TITLE	End sequencing of clones from a full length enriched, normalized JEC21 cDNA library
JOURNAL	unpublished (2003)
COMMENT	Other ESTs: CCAP099TR Contact: Brendan Loftus
	TIGR
	Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA Tel: 301-838-3543 Fax: 301-838-0208 Email: crypt@tigr.org Seq primer: TF
FEATURES	
source	Location/Qualifiers 1..890

FEATURES	source	Location/Qualifiers
1. .890		
/organism="Corynebacterium neoformans var. neoformans"		
/mol_type="rRNA"		
/strain="JEC21"		
/db_xref="taxon:40410"		
/clone="GCAPB09"		
/clone_11b="C. neoformans strain JEC21"		
/note="Vector: pCMVSPORT6; Site 1: NotI. EcoRV. The full		

length, normalized library was prepared from a variety of conditions using RNA provided by Joseph Heltman and Jennifer Lodge"

Alignment Scores:	
Pred. No.:	1,076-36
Score:	500.00
Percent Similarity:	55.36%
Best Local Similarity:	41.79%
Query Match:	13.99%
DB:	7
length:	890
Matches:	117
Conservative:	38
Mismatches:	83
Indels:	42
Gaps:	6

US-10-089-514-2 (1-686) x CF703510 (1-890)

Oy		446	ProthrGluAlaThrIleLeuProLeuTyrSerAlaLeuAGAlaIleSerProValPro	465
Db		888	CCTTCATCTTTGCATCCTTTGGCTGTACTTGACCACTTCGGAAATTCAACCCGCCCTAT	829
Oy		466	TyrGlYAlaLeuLeuGluPheProGluLeuSer	476
Db		828	TACTCTACCTATATAGATTTCGCCGCCCTTCMAA CTTCTCATATACATGGACAGCTCAGA	769
Oy		477	-----ValLeuSerAlaSerProGluArgPheLeuThrIleGlYAlaAsp	491
Db		768	ACACAGGGATACACCATCTCTCAAGTCTCTGAACGGTCTTGAAAATCAACCTCTCC	709
Oy		492	GlyGlYAlaGluSerLysProIleLysGlYThrArgProArg	505
Db		708	AGACAAGTGGANGATGATGCCCATCAGAGGAACGAGACGATCAAGAAACCAATGT	649
Oy		506	-----GlyGlYThr	508
Db		648	ATATGTCAGCTCGGCATCGGGTGTGGGGGAAAAAACCAAGCCTGCTGGTAGTCAAACAG	589
Oy		509	-----AlaGluGluAspGluArgLeuArgAlaAspLeuAlaGlyArgGluLysAspArg	526
Db		588	GAAAGGAGAAAGGGAGATGAGCGGAAGAGGGAGAAATTGCTTGAAGATGCAMAAGACGG	529
Oy		527	AlaGluAsnLeuMetIleValAspLeuValArgAsnAspLeuAsnSerValCysAlaIle	546
Db		528	GCGAAAAACCTCATGATGTTGATCTTAATCCGATTCGATTACTCTCATGTTGTAATCCCG	469
Oy		547	GlySerValHisValProArgLeuPheGluValGluThrTyralaProValHisGlnLeu	566
Db		468	TCAACAGCTCACCGTTCCAATAACTTAAGCTCTTAATCTAATGG--GTGCATAACCTA	412
Oy		567	ValSerThrIleArgGlyArgLeuArgProGlyThrSerThrAlaAlaCysValArgAla	586
Db		411	GTGACAAACAGTGCAGAGCAGTTGGCAGATATGTTGGAGTGGAGACGTCAAAAGA	352
Oy		587	AlaPheProGlyGlySerMetThrGlyAlaProLysLysArgThrmGluIleLeuArg	606
Db		351	TGTTTTCCCCTGGTTCATGACGGGGGACC CGAAGTTGAGGTCACTACACTTTTGAT	292
Oy		607	ArgLeuGluGluGlyPro--ArgGlyValTyrSerGlyAlaLeuGlyTyrPheAlaLeu	625
Db		291	GAGTTTGAGAAACATCAGCAAGGGGGGTATACCTCGGTCTTTGGTTACTTTTCTGTCC	232
Oy		626	SerGlyAlaAlaAspLeuSerIleValIleArgThrIleValLeuAlaAspGlyGlnAla	645
Db		231	GATGAGTCAACAGATCTCAGTGTGTATCTCGTACGATTCGATGTCGAGATPAATCAATTA	172
Oy		646	GluPheGlyAlaGlyGlyAlaIleValSerLeuSerArgGlnGluGluPheThrGlu	665
Db		171	AGTATCGGGCTGGAGGTCCATCACTCGCTATTCGATTAAGGGAGAAATGGAGTAG	112
Oy		666	ThrValValLysAlaArgAlaMetValThrAlaLeu--AspGlySerAlaValAlaGly	684
Db		111	GTGTTCACCAAGGGAAGTCTGTAGTTGGAAGAATATACAGATGAGAAATGACATGCTTGGC	52

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:00:45 ; Search time 1433.22 Seconds
(without alignments)
10548.289 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312
Sequence: 1 atgaccgagcagcagcagc.....cggcgcgagccgcgcgtga 312

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:*
1: gb da:*
2: gb hlg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_dr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	312	100.0	312	6	BD178314 Transform
2	312	100.0	312	6	BD093915 Transform
3	312	100.0	5251	1	AB116234 Streptomy
4	296.4	95.0	14159	1	AF262220 Streptomy
5	135.6	43.5	387	6	A48325 Sequence 3
6	135.6	43.5	2888	6	AR198355 Sequence 1
7	135.6	43.5	2888	6	AR198353 Sequence 1
8	135.6	43.5	2888	6	AR198353 Sequence 1
9	135.6	43.5	2888	6	AR198353 Sequence 1
10	64	19.5	13635	1	AE005068 Halobacte
11	60.8	19.5	13635	1	AE005068 Halobacte
12	59.4	19.0	1964	8	AK121213 Oryza sat
13	58.2	18.7	16905	1	AE005050 Halobacte
14	55.8	17.9	110000	1	AP006618_43 Continuation (29 o
15	55.8	17.9	110000	1	AP006618_43 Continuation (29 o
16	55.4	17.8	300800	1	SC0939112 Streptomy
17	55.2	17.7	346259	1	SC0939112 Streptomy
18	55	17.6	132544	1	AF521085 Streptomy
19	55	17.6	299300	1	AP005026 Streptomy

20	54.8	17.6	1458	6	AX573996 Sequence
21	54.8	17.6	45055	6	AX574197 Sequence
22	54.8	17.6	300800	1	AP005036 Streptomy
23	54.8	17.6	349028	1	BX640413 Bordetelli
24	54.6	17.5	1349	4	BOV72TRANS
25	54.6	17.5	73882	6	CQ870950 Sequence
26	54.6	17.5	89713	1	AJ605139 Actinopla
27	54.6	17.5	89976	1	AJ632270 Actinopla
28	54.6	17.5	110000	1	AP006618_28 Continuation (29 o
29	54.4	17.4	110000	1	AP006618_43 Continuation (29 o
30	54.4	17.4	17050	1	AL646081 Ralstonia
31	54.4	17.4	321250	1	SC0939111 Streptomy
32	54.4	17.4	343473	1	BX640451 Bordetelli
33	54.4	17.4	346259	1	BX640435 Bordetelli
34	54.4	17.4	349726	1	BX640421 Bordetelli
35	54	17.3	1208	6	A37834 Sequence 4
36	54	17.3	1208	6	AR069872 Sequence
37	54	17.3	1208	6	AR099268 Sequence
38	54	17.3	1208	6	AR124153 Sequence
39	54	17.3	1208	6	AR442760 Sequence
40	54	17.3	5392	6	A37831 Sequence 1
41	54	17.3	5392	6	AR069869 Sequence
42	54	17.3	5392	6	AR099265 Sequence
43	54	17.3	5392	6	AR124150 Sequence
44	54	17.3	5392	6	AR442757 Sequence
45	54	17.3	190050	1	AL646080 Ralstonia

ALIGNMENTS

RESULT 1
BD178314
LOCUS
DEFINITION
BD178314 Transformant producing pF1022 substance and process for producing the same and novel biosynthetic gene.
ACCESSION
BD178314
VERSION
BD178314.1 GI:30015579
KEYWORDS
WO 02077244-A/2.
SOURCE
Streptomyces venezuelae
ORGANISM
Streptomyces venezuelae
Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
REFERENCE
1 (bases 1 to 312)
Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.
Transformant producing pF1022 substance and process for producing the same and novel biosynthetic gene
Patent: WO 02077244-A 2 03-OCT-2002;
MORIYU SRIKA KAISHA LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI

COMMENT

OS Streptomyces venezuelae
PN WO 02077244-A/2
PD 03-OCT-2002
PR 22-MAR-2002 WO 2002JP002782
PT KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI
PI C12N15/61, C12N1/15, C12P17/14// (C12N1/15, C12R1:645), (C12P17/14, C12R1:645)
PC Transformant producing pF1022 substance and process for CC
CC producing the same
CC and novel biosynthetic gene
FH Key location/Qualifiers
FT CDS location/Qualifiers (1) . (309).

FEATURES

source
1. 312
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"

ORIGIN

Query Match .100.0%; Score 312; DB 6; Length 312;

Best Local Similarity 100.0%; Pred. No. 4.1e-33;
Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

[illegible]

Qy 61 CTCTGGACACGGTGGCGCCCGCATGACCTCGGTGTCGCCGATCGCGCGGTACAAAGTCC 12
|||
Db 61 CTCTGGACACGGTGGCGCCCGCATGACCTCGGTGTCGCCGATCGCGCGGTACAAAGTCC 12

Qy 121 CGGCACGGCGTCCCGATGATCAGCCCGGCCGGTCAGCCTGGTCAAGGACAGGGCCGCC 18

Db 121 CGGCACGGCGTCCCGATGATCAGCCCGGCCGGTCAGCCTGGTCAAGGACAGGGCCGCC 18

Qy 181 CGCTACGCCCGGACCAAGGCTCGAGCAATCGTTCCTGGTAACCTCTACGACGATGC 24
 |||||
 Db 181 CGCTACGCCCGGACCAAGGCTCGAGCAATCGTTCCTGGTAACCTCTACGACGATGC 24

QY 241 ATCAGCGAGATGTGCGCGCTGAGGAGACTGGATGAGCGGCGGAG 30
 |||||
 Db 241 ATCAGCGAGATGTGCGCGCTGAGGAGACTGGATGAGCGGCGGAG 30

QY	301	GACCGCGGTGA	312
Db	301	GACCGCGGTGA	312

RESULT 2				
BD093915				
LOCUS				
BD093915	312 bp	DNA	linear	PAT 27-AUG-20

ACCESSION
VERSION
BD093915.1
GI:22639503

SOURCE ORGANISM	REFERENCE
<i>Streptomyces venezuelae</i>	1
<i>Streptomyces venezuelae</i>	2
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	3

REFERENCE
AUTHORS
1 (bases 1 to 312)
Yanai, K., Okakura, K., Yasuda, S., Watanabe, M., Miyamoto, K., Mido, N.
and Murakami, T.

functional group(s) and novel biosynthesis genes
Patent: WO 0123542-A 2 05-APR-2001;
MEIJI SEIKA KAISHA LTD KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA,

COMMENT	OS
	Streptomyces venezuelae
	WO 0123542-A/2
	PN
	05-APR-2001
	PD
	28 SEP 2000 WO 2000/060373

PR 29-SEP-1999 JP 99P 276314
PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI
PI MIYAMOTO,
PI NAOKI MITO, TAKEGUCHI MITSURUMI

PC	C12N15/09, C12N5/10, C12P21/02, C07K11/00//C12P21/02, C12R1:645
CC	
PH	
Key	Location/Qualifiers
FT	(1)
CDS	(309)

```

FEATURES
source      Location/Qualifiers
1. 312
/organism="Streptomyces venezuelae"
/mol type="genomic DNA"

```

Query Match	Score	DB 6	length
100.0%	312	312	312

```

0Y      1 ATGACCGAGCAGACGAGCTTCAGCGGCTCGACGCTTCGACGCGGAGC 60
      best local similarity 100.0%;  Freq. NO. 4, 1e-33;
      Matches 312;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps

```

D**b** 1 ATGACCGAGCAAGAGAGCTGCAGCGGCTGCGCGGGAGCTCGACGCCCTCGACGGGACG 60

61 CTCCTGGACACGGTCGGCGCGCATCGACCTCGGTGTCGCGCATCGCGCGGTACAAGTCC 120

121 CGGACGGGCTCCCGATGATCAGCCCGCGGGTCACTGGTCAAGGACAGGGCGGCC 18

Db

181 CGCTACGCCCGACCAAGGCTCGACGAATGTTCTTGTAACCTTACGACGTGATC 241

241 ATACGAGATGTGCCGCTGAGACTGTGATGAGCCGAGAGCCTGACGCGGAG 30

Db 301 GACCGGGCGTGA 312

LOCUS	AB116234	5251 bp	DNA	linear	BCT 03-JUL-20
DEFINITION	Streptomyces venezuelae pApC, pApC, pApB, ORFV genes for				
RESULT 3	AB116234				

4-aminino-4-deoxycycloisomate synthase, 4-aminino-5-deoxypropanate dehydrogenase, 4-aminino-4-deoxycycloisomate mutase, hypothetical protein, complete cds.
AB116234
ACCESSION

VERSION	AB110234.1	GI:147040005
KEYWORDS		
SOURCE	Streptomyces venezuelae	
ORGANISM	Streptomyces venezuelae	

REFERENCE
AUTHORS

1
Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and

TITLE
Para-position derivatives of fungal anthelmintic cyclodepsipeptide
engineered with *Streptomyces venezuelae* antibiotic biosynthetic
genes

PUBMED	15184904
REFERENCE	2 (bases 1 to 5251)
AUTHORS	Yanai, K., Sumida, N., Okakura, K., Moriya, T., Watanabe, M. and

TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Koji Yanai, Meiji Seika Kaisha, Ltd.,
Microbiological Resources and Technology Laboratories, 788 Kayama
2-1-1, Kayama, Osaka 591-8501, Japan

(E-mail: kot_i_yanaimeiji.co.jp, Tel: 81-465-37-5106,
Fax: 81-465-37-6397)

```

/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
seq_ 3711

```

```

CDS      /gene="papa"
         651..2711
         /gene="papa"
         /codon_start=1

```

```
/evidence=experimental
/transl_table=11
/product="4-amino-4-deoxychorismate synthase"
/protein_id="BAD21140.1"
```

```

/db_xref="GI:47846866"
/translation="MRTLLIDNYSFTNLFQYIGSATGQPPVVPENDADMSRLPV
FDALVSPGSGPDRERDGISRAITDSGLPVLGCLGHGSIQLFGGTGLAPE
HGRVSEVATGEDVFERGLPSPTFVRVHSLAATDLPDELPLAMSDGVVMGLRHR

```

FUMGVQEHFESLSUFGKBIWMPKULALHHRKAKKHGAUSFIELHVKKVDVLEPDA


```
/product="Cm1H"
/protein_id="AAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITDLPFRHGRGRLARTAGRTWTVEELDRVTSALARRID
AECPRARVIVAGEHTAEAVYVWALAMRSHAVTTPNRCPLPDRPREFRVAADALIV
CERREALVNGEKAGLALYAGVGMPTDPPADGTADEPARSAVSIPTSGTGDP
PKLVGHGGLNLCSRLRLDITPDQDLVHSAISFASVSEILGTLYAGATLVVP
RQDAWLGSVSHLAHGCDLMLSPSVARLDEPARSIRVSEFCGALGGEYDK
LAARVFNAYGPTAEVCSFLAELTSTPISGTSPVGFRAVYRDPDSDGHAATGE
LVVGGVALGVAGSGSPAENEVGVDSAPVATGVDSIDGELTYGRIDEQIKR
LGRVNLAVGSGTSLRHLGREVALVRODATTLLVTPADDEAPRESLMARIDVYME
APRVLVNDALPTTSGKTRDRALRLILSPACAPRGCTDGEDAALRVLDVYAVL
GQETGPETSI FDAQSSSLAMTQIVKLSAYGEAEVAAADYDFAPAFRLHNG
EAVAPAESAVDTLLRREVERTERDRLAELRLRDTHEBPVPGAADGDRVLLTGASG
IGHVAIDRLLAAGRPLVVTGDPDGLTGHAATRPFORAADYARVRAISYAELEKRV
RRGPVDAVVGCGOVNHLPLDLSGSTRVTLVRAAALGARSAFPLSAASAG
ADPLISAATLTAVGDPYRSKLSSEYVNTTLAVGCANVTRPGVYGRPREDRHL
KDDWFTALLETARRVAMPRLSGHVPCVDITLADTLGPRDANPGTAGASTPDR
SAVVHRTYALDELRLHGTLEADVLAAPAFERVRDGEVAPLLAAQALSGPW
PSAHRVDHDIQLRLIGTPDTPADGDRPERTGTSRGPERRRRPRADTALTGTEP"
```

Query Match 95.0%; Score 296.4; DB 1; Length 14159;
Best Local Similarity 98.4%; Pred. No. 2e-31; Mismatches 1; Indels 4; Gaps 1;
Matches 311; Conservative 0;

misc_feature /note="Region: adenylation domain A1"
misc_feature /note="Region: adenylation domain A2"
misc_feature /note="Region: adenylation domain A3"
misc_feature /note="Region: adenylation domain A4"
misc_feature /note="Region: adenylation domain A5"
misc_feature /note="Region: adenylation domain A6"
misc_feature /note="Region: adenylation domain A7"
misc_feature /note="Region: adenylation domain A8"
misc_feature /note="Region: adenylation domain A9"
misc_feature /note="Region: adenylation domain A10"
misc_feature /note="Region: thiolation domain"
misc_feature /note="Region: NADH binding site"

ORIGIN

Query Match 95.0%; Score 296.4; DB 1; Length 14159;
Best Local Similarity 98.4%; Pred. No. 2e-31; Mismatches 1; Indels 4; Gaps 1;
Matches 311; Conservative 0;

1 ATGACCGAGCAGAAAGAGTGA----GGCGTGGCGCGGAGAGCTGACGCGCTCGACGG 56
11142 ATGACCGAGCAGAAAGAGTGA----GGCGTGGCGCGGAGAGCTGACGCGCTCGACGG 11201

57 GACGCTCTTGAGACAGGTGGCGCGCGCATGACCTTCGATGTCGATCGCGGAGTACA 116
11202 GACGCTCTTGAGACAGGTGGCGCGCGCATGACCTTCGATGTCGATCGCGGAGTACA 11261

117 GTCCCGGAGCAGCGGTCCCATGATGACAGCCCGCGGATGACCTTGTCAAGACAGGGC 176
11262 GTCCCGGAGCAGCGGTCCCATGATGACAGCCCGCGGATGACCTTGTCAAGACAGGGC 11321

177 CGCGCGCTACGCGCGGAGCAGCGGCTCGAGATGTTCTGTGTAACCTTACAGACGT 236
11322 CGCGCGCTACGCGCGGAGCAGCGGCTCGAGATGTTCTGTGTAACCTTACAGACGT 11381

237 GATCATCACGAGATGTCGCGCTCGAGGACCTGTGATGAGCCGGAGAGACCTTCAGCGC 296
11382 GATCATCACGAGATGTCGCGCTCGAGGACCTGTGATGAGCCGGAGAGACCTTCAGCGC 11441

297 CGAGGACCGGCGGTGA 312
11442 CGAGGACCGGCGGTGA 11457

RESULT 5
A48325
LOCUS A48325 387 bp DNA linear PAT 07-MAR-1997
DEFINITION Sequence 3 from Patent WO9601901.
ACCESSION A48325
VERSION A48325.1 GI:202118
KEYWORDS Streptomyces pristinaespiralis
SOURCE Streptomyces pristinaespiralis
ORGANISM Streptomyces pristinaespiralis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 387)
Auteurs Blanc V., Thibaut D., Bamas-Jacques N., Blanche F., Crouzet J.,
Barriere J., Debussche L., Famechon A., Paris J. and
Dutruc-Rosset G.
TITLE Streptogramins AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS
JOURNAL Patent: WO 9601901-A 3 25-JAN-1996;
COMMENT Other publication AU 2891295 960209
Other publication FR 2722210 960112.
FEATURES
source Location/Qualifiers
1..387
/organism="Streptomyces pristinaespiralis"
/mol_type="unassigned DNA"
/db_xref="taxon:38300"

ORIGIN

Query Match 43.5%; Score 135.6; DB 6; Length 387;
Best Local Similarity 69.8%; Pred. No. 2.5e-09;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

14 ACCGAGTGAAGCGGCTGGCGCGGAGAGCTGAGCCCTCGACGGGACGCTCTGACACGG 73
59 ACCCTTCGACGGGCTGGCGCGGAGAGCTGAGCCCTCGAGACGG 118
74 TCGGCGCGCATGCACTCGGATGCCGATCGCGGATCAAGTCCCGCACGGCGTCC 133
119 TCCGACACGCTGTGACATCTGCTCGGATGCGGAGTCAAGGCGCTCCACAGTGC 178
134 CGATGATGAGCCCGCGGCTGACGCTGTCAGAGCAGGCGCGGCTGACGCGCG 193
179 CGATGATGAGCCCGCACCGGATGCGGAGTCCAGGTCACGSCAACCGCGCGCTGACGCGCG 238
194 ACCAGCGCTCGACGCAATGTTCTGATGACCTTACGACGTCATGATGATGATGAT 253
239 ACCAGCGCATGACCGCGCTTCTGCGGACCTGTGACGACGATCATACGAGACCT 298
254 GCCGCTCGAGGACCTGTGAT 275
299 GCCGCTCGAGGACGATGAT 320

RESULT 6
A198355
LOCUS A198355 387 bp DNA linear PAT 20-APR-2002
DEFINITION Sequence 4 from patent US 6352839.
ACCESSION A198355
VERSION A198355.1 GI:20248204
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 387)
Auteurs Blanc V., Thibaut D., Bamas-Jacques N., Blanche F., Crouzet J.,
Barriere J., Debussche L., Famechon A., Paris J., M. and
Dutruc-Rosset G.
TITLE Streptogramins for preparing same by mutasynthesis
JOURNAL Patent: US 6352839-A 4 05-MAR-2002;
FEATURES
source Location/Qualifiers
1..387
/organism="unknown"

ORIGIN /mol_type="unassigned DNA"

Query Match 43.5%; Score 135.6; DB 6; Length 387;
Best Local Similarity 69.8%; Pred. No. 2.5e-09;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGACGAGCGCTGCGCGGAGCTCGACGCGCTTCGACGGAGCGCTCTGGACACGG 73
DB 59 ACCCCCTCGACGGCGCTGCGCGGCGCCCTGAGACGCGCGGAGCGCCGCTTCTGGACGCGC 118
QY 74 TCGGGCGCGCGCATGACCTGCTGCTGCGCATTCGCGCGTACAAAGTCCCGGACGCGCTCC 133
DB 119 TCGCACACGCGCTGACATCTGCTGCGCATTCGCGCGAGTCAAGCGCTTCACACGAGGTG 178
QY 134 CGATGATGACGCGCGCGCGCGCTGAGCTGCTCAAGACAGGGCGCGCGCTTACGCGCGC 193
DB 179 CGATGATGACGCGCGCGCGCGCGATCGCCAGGTCCACGCAAGCGCGCGCTTACGCGCGC 238
QY 194 ACCACGCGCTCGACGAACTGCTTCTGCTGTAACCTCTACGAGTATCATACGAGATGT 253
DB 239 ACCACGCGCATGACCGCGCTTCTCTGCGCACCTCTGACAGATATCATACGAGACT 298
QY 254 GCCGCTCGAGGACCTGTGTAT 275
DB 299 GCCGCTCGAGGACGAGTGTAT 320

RESULT 7
A48323 2888 bp DNA linear PAT 07-MAR-1997
LOCUS Sequence 1 from Patent WO9601901.
ACCESSION A48323
VERSION A48323.1 GI:2302116
KEYWORDS Streptomyces pristinaespiralis
SOURCE Streptomyces pristinaespiralis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1 (bases 1 to 2888)
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,
Dutric-Rosset,G.
TITLE STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS
JOURNAL RHONE-POULENC KORER SA (FR)
COMMENT Patent: WO 9601901-A 1 25-JAN-1996;
Other publication AU 2891295 960209
Other publication FR 2722210 960112.
FEATURES
source 1..2888
/organism="Streptomyces pristinaespiralis"
/mol_type="unassigned DNA"
/db_xref="taxon:38300"

ORIGIN

Query Match 43.5%; Score 135.6; DB 6; Length 2888;
Best Local Similarity 69.8%; Pred. No. 1.5e-09;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGACGAGCGCTGCGCGGAGCTCGACGCGCTTCGACGGAGCGCTCTGGACACGG 73
DB 1931 ACCCCCTCGACGGCGCTGCGCGGCGCCCTGAGACGCGCGGAGCGCCGCTTCTGGACGCGC 1990
QY 74 TCGGGCGCGCGCATGACCTGCTGCTGCGCATTCGCGCGTACAAAGTCCCGGACGCGCTCC 133
DB 1991 TCGGCACACGCGCTGACATCTGCTGCGCATTCGCGCGAGTCAAGCGCGCTTCACACGAGTGC 2050
QY 134 CGATGATGACGCGCGCGCGCGCTGAGCTGCTCAAGACAGGGCGCGCGCTTACGCGCGC 193
DB 2051 CGATGATGACGCGCGCGCGCGCGAGTGCACGCGCAAGCGCGCGCTTACGCGCGC 2110
QY 194 ACCACGCGCTCGACGAACTGCTTCTGCTGTAACCTCTACGAGTATCATACGAGATGT 253
DB 299 GCCGCTCGAGGACGAGTGTAT 320

DB 2111 ACCACGCGCATGACCGCGCGCTTCTGCGGACCGCTGTACGACAGATCATACGAGACT 2170

QY 254 GCCGCTCGAGGACGAGTGTAT 275
DB 2171 GCCGCTCGAGGACGAGTGTAT 2192

RESULT 8
AR198353 2888 bp DNA linear PAT 20-APR-2002
LOCUS Sequence 1 from patent US 6352839.
ACCESSION AR198353
VERSION AR198353.1 GI:20248202
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 2888)
AUTHORS Blanc,V., Thibaut,D., Bamas-Jacques,N., Blanche,F., Crouzet,J.,
Barriere,J.-C., Debussche,L., Famechon,A., Paris,J.-M. and
Dutric-Rosset,G.
TITLE Streptogramins for preparing same by mutasynthesis
JOURNAL Patent: US 6352839-A 1 05-MAR-2002;
FEATURES
source 1..2888
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 43.5%; Score 135.6; DB 6; Length 2888;
Best Local Similarity 69.8%; Pred. No. 1.5e-09;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACAGAGCTGACGAGCGCTGCGCGGAGCTCGACGCGCTTCGACGGAGCGCTCTGGACACGG 73
DB 1931 ACCCCCTCGACGGCGCTGCGCGGCGCCCTGAGACGCGCGGAGCGCCGCTTCTGGACGCGC 1990
QY 74 TCGGGCGCGCGCATGACCTGCTGCTGCGCATTCGCGCGTACAAAGTCCCGGACGCGCTCC 133
DB 1991 TCGGCACACGCGCTGACATCTGCTGCGCATTCGCGCGAGTCAAGCGCGCTTCACACGAGTGC 2050
QY 134 CGATGATGACGCGCGCGCGCGCTGAGCTGCTCAAGACAGGGCGCGCGCTTACGCGCGC 193
DB 2051 CGATGATGACGCGCGCGCGCGAGTGCACGCGCAAGCGCGCGCTTACGCGCGC 2110
QY 194 ACCACGCGCTCGACGAACTGCTTCTGCTGTAACCTCTACGAGTATCATACGAGATGT 253
DB 2111 ACCACGCGCATGACCGCGCGCTTCTGCGGACCGCTGTACGACAGATCATACGAGACT 2170
QY 254 GCCGCTCGAGGACGAGTGTAT 275
DB 2171 GCCGCTCGAGGACGAGTGTAT 2192

RESULT 9
SPU60417 4740 bp DNA linear BCT 07-MAR-1997
LOCUS SPU60417
DEFINITION Streptomyces pristinaespiralis 4-dimethylamino-L-phenylalanine
precursor biosynthesis (papA, papC, papB, papM) genes, complete
cds.
ACCESSION U60417
VERSION U60417.1 GI:1575335
KEYWORDS Streptomyces pristinaespiralis
SOURCE Streptomyces pristinaespiralis
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
REFERENCE 1 (bases 1 to 4740)
AUTHORS Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Zagorec,M.,
Thibaut,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and
Schleutinger,J.
TITLE Identification and analysis of genes from Streptomyces
pristinaespiralis encoding enzymes involved in the biosynthesis of

JOURNAL Mol. Microbiol. 23 (2), 191-202 (1997)
MEDLINE 97197164
PUBMED 5044253
REFERENCE 2 (bases 1 to 4740)
AUTHORS Blanc, V., Gil, P., Bamas-Jacques, N., Lorenzon, S., Schlenninger, J.,
Bisch, D., Blanche, F., Debussche, L., Crouzet, J. and Thibaut, D.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc
Rorer, 13 quai Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403,
France

FEATURES
source location/Qualifiers
1..4740
/organism="Streptomyces pristinaespiralis"
/mol_type="genomic DNA"
/db_xref="taxon:38300"
68..2227
/gene="papa"
68..2227
/gene="papa"
/function="p-aminobenzoate synthase"
/codon_start=1
/transl_table=11
/product="Papa"
/protein_id="AAC44866.1"
/db_xref="GI:1575336"
/translation="MRTVRLIDNYDSFTYNNLFOMLAEVNGAPLVVRNDRTWQA
LAPDFDNVVSFPGPATDIDLGSRVITEMDIPLDGCHOLGICLAQAAVH
APBPHGRSDIRHDOGLEFANTPSPLTVRVHSLVRG.PADLRPATRAGDGLAV
AHRHLPRFGVQFHPESI SEHGRMLANPRDLSRAAGHPHTTERIPAPAPAPAP
APAPAPAPVGEYRLHREYACVPDADAAFTALFADAPAFWLDSSRVEGLARFTFL
GAPAPGLGEITVDADRAVAVDGGSGGEGTRPGTLFDLHEBLARALPATGLPEF
NLGYVYLGEYTKADSGEDAHRGELPDGAFMADRLALDHEGABMALTAGSTERP
ATAPAEKMLTDATLATTAPAPFLLPDDQPLADYHRSGLPREYLVESGRL
ITGGEYVCLTNMLRVPGRIDPLTVRLRTVSPAPAYIQPGATYSSPSEEL
RIGADMAESKPIKTRPRGAGAPADAAVYASIAAKRSENMIMVDYRNDLGCVC
DISGVVPGLEVEYTAIVHQLVSTVGRILAADVSRPRAVRAFPDGSMTGAKVETM
QFIDRLKRGVYSGALGFALSGAADISIVRTIVATEEAATIGVGAAVVALSDPD
DEVREMLKQITLALRQAGATASDBELLASGR"
/complement (2486..3376)
/gene="papC"
complement (2486..3376)
/gene="papC"
/function="dehydrogenase"
/codon_start=1
/transl_table=11
/product="PapC"
/protein_id="AAC44867.1"
/db_xref="GI:1575337"
/translation="MRGGSYFGRGVVVGAGAVGRNFHMLVRSVAVTWLDDVAGAGA
ADGVRVAGDVRPRPEAVNALAADVVVLAPEPVAMEAVEVLAVMPGAVIADTL
SVKSRVAGRLREAPGLQAVGNPMFAPSLGQGRVAAVVTVDGPAVLAIDLTL
GAIVMPAPRRHDELTAQQAATHAVALFGCLGSLAVVVGALRPSAPPHAMAL
LARIAGTPEYVTDIOANFGAPAPAAQALGRGLVRLGQAVRDEETFAALFAELKGV
LGHGAELERLCRMFTALH"
3413..3802
/gene="papB"
3413..3802
/gene="papB"
/function="mutase"
/codon_start=1
/transl_table=11
/product="PapB"
/protein_id="AAC44868.1"
/db_xref="GI:1575338"
/translation="MTPPAIPAPPAATGPAATDIDALRARDADADALLDAVRRL
DLCIRIGERYRLHQVPMQPHRIAYVHANAARVAADHGIDPAFLRTLDTITETCRLL
BEMWISGAGAPVPTPVHASASARGAVS"
3799..4677
/gene="pamM"
3799..4677
/gene="pamM"
/function="N-methylase"

ORIGIN

Query Match 43.5%; Score 135.6; DB 1; Length 4740;
Best Local Similarity 69.8%; Pred. No. 1.3e-09;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACGAGCTGCAGCGCGCTGCGCGGAGACTGCAGCGCCCTCGACGGAGCGCTCTGACACGG 73
DB 3471 ACCCCCTGCAGCGCGCTGCGCGCGCGCTGCAGCGCCCGACCGCGCTCTGACCGCG 3530
QY 74 TGGCGGCGCGCATCGACTCGGTGTCGGATCGCGGCTACAGAGTCCCGGACAGGGCTCC 133
DB 3531 TCCGACACGCGCTGCAGCATCTGCTCGGATCGGCGGACAGTCCGCTTCACACAGGTGC 3590
QY 134 CGATGATGAGAGCCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193
DB 3591 CGATGATGAGAGCCCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3650
QY 194 ACCACGCGCTCGACGAATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
DB 3651 ACCACGCGCATCGACCGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3710
QY 254 GCCGCGTCGAGCGACTGCTGAT 275
DB 3711 GCCGCGTCGAGCGACTGCTGAT 3732

RESULT 10

AF263011S2/c

LOCUS AF263011S2 10029 bp DNA linear BCT 07-JUN-2000
DEFINITION Streptomyces griseus subsp. griseus macroretroilide biosynthesis
gene cluster, downstream region.

ACCESSION AF263012
VERSION AF263012.1 GI:8896099

KEYWORDS

2 of 2
Streptomyces griseus subsp. griseus
Streptomyces griseus subsp. griseus
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.

SOURCE

2 of 2
Streptomyces griseus subsp. griseus
Genetic localization and molecular characterization of the nons
gene required for macroretroilide biosynthesis in Streptomyces

AUTHORS

Smith, W.C., Xiang, L. and Shen, B.
Genetic localization and molecular characterization of the nons
gene required for macroretroilide biosynthesis in Streptomyces
griseus DSM40695

TITLE

Antimicrob. Agents Chemother. 44 (7), 1809-1817 (2000)
10858335

JOURNAL

Submitted (02-MAY-2000) Chemistry, University of California at
Davis, One Shields Ave., Davis, CA 95616, USA
Location/Qualifiers
1..10029
/organism="Streptomyces griseus subsp. griseus"
/mol_type="genomic DNA"
/strain="DSM40695"
/sub_species="griseus"
/db_xref="taxon:67263"
205..990
/gene="nonH"

CDS
205.. .990
/gene="nonH"
/note="similar to fixa"
/codon_start=1
/transl_table=11
/product="flavoprotein reductase nonH"
/protein_id="AA081240.1"
/db_xref="GI:8896112"
/translation="MSRIYCVKYYVDATGDRPADDTLDEEDVDGLSELDEVAV
EOLQIADGADDEITVTVGPEDAKALRKALSMGADKAHVEDDLHSDVWGTS
VLAAVEKAGYDLVTICMASTIDWGLPALLAEIRLGPQVTLISVSGRGVNGTGR
LDCATSVDESLASIPAVSVTDGSEAPSPFGKIMAAKKPVESLDDLDGLDADEVG
LAGMTVDATSPARTACTIVDEBEGGRLAGLAGQKFI"
1051..2013
/note="orf26; similar to fixb"
/codon_start=1
/transl_table=11
/product="flavoprotein reductase"
/protein_id="AA081241.1"
/db_xref="GI:8896113"
/translation="MAEVLVVDHVGCAVAKPTLELTTLARRVDPVAVVLGAGAGA
APVDEHGAVKVLTSDAPEFADYLVKPAALQAAFAVSPVAVLVSSAEKEIAR
LALRVSGIITDITDEAGDQGVATQAPALASFTTSRSGVPIVTVNSAPVPEP
APAGAVALPVEFGEATGTKTSRPRESTGRPELTETAIIVSGRGVNGAENFPL
IEALADSLGAAGVAGRAVDAGWPHTSOYGOTGKSVSPOLYIASGISGAIGHAGMO
TSKTIYAINKDAAPIFEIYDGVGDLFAVYQLTPEINTRK"
complement (2090..3481)
/note="orf27"
/codon_start=1
/transl_table=11
/product="amino acid monooxygenase-like protein"
/protein_id="AA081242.1"
/db_xref="GI:8896114"
/translation="MTPGPRTPPLDLLGVAGPNNLSLALADGVPGLRTAFHEHQR
FHHMGLIGATLOVFLADLVSLVEPTSPWLYNLKIRRLFPYFAEHFHIRT
EFNNYLRWVSTELPATRFGRVDVTAMDPODAVAVEFTLGDGELTGLTHARL
ALGVGTAPHPEPRLVODPAALVHASAYLSORNLAAHITVYSGSGSABVIL
DLARSREGAGLTWLTARTAPAPMEYSKIGLEQFTDYTRYHGLPQARLDLPRO
WQILKVGSGTLDGIDELYRSIDGSMVPVTLPGLEVGAATEGRVLELVENQ
QEARGLVTVADVLATGVRERPVDLAALADPYI VREOTGRPOVDAQRLVLAPEL
SVFVONARHTHGTGTPDLGLAAMRSAYIVNTLTGKEFYLPERTAFPTTGLGARBD
DRDTASRAEERR"
complement (3552..5129)
/note="orf28"
/codon_start=1
/transl_table=11
/product="amino acid decarboxylase-like protein"
/protein_id="AA081243.1"
/db_xref="GI:8896115"
/translation="MSDGTPTPIPGPACPRGCVAGAGHGRRLAULTAIVLALDEMA
PARGRLPAGGEVVALSTLARCAPVLEEGVPEALGIVRAVAGADPADPACV
AHLACPLAAVAVADLADALNPSMDSMDPAGVIEELTALALHPTAPADPA
LITSGGTESNLVALLARERAVGTLRVGVANSHRAAMMLGLPAPVVAACRG
RIDPALDDLASLAGAPLTVATAGTDDGRIDPPEIRAVRERHARLHVAAVAGG
PLFSERLAFLAGLEHAASVTDLHGLGQVPAAGVLAADAGMLAPLSLADYANA
DDTBAGLPDLIGRSITTRRPDLAKAVTFRALGRKGLSELVEHCRTKTAHERGAVD
TRPGLRRRPEDIGISTVLFRTVADALPDAGAAVVAAVRITLLAERAVLGAQMED
GDGRRRLMKATLHPDAAAELPLDLIVTAADRAATEADGVVGTADADRAATE
APGAEITDAPAAERLSP"
complement (5438..6403)
/note="orf29"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AA081244.1"
/db_xref="GI:8896116"
/translation="MREVTCHVEACQRLDALSGISGRITGRHMKRYDDASERLRE
TABLDLHVAVAADLDDVARSALRTAASCHLGWMSGCFPDGQELHPLIGIT
LTSBEDLAFGVVRRGGQAPSAGTWDAFAVLSVSGVWDMQVITGLINDVAPAH
DGVYSLDASDPTDLAAMDALCPYLAEAGHQPRLMPTVPLRRPADLIALAAAL
DASGALTPDQRLVLDDDOAAFEELASRLVGRHVSVPDPAPRTLLPLHALAAL
LAVOAHGMEIGLSGVLPPELLGTPPALRAAGSRTNNLGWYAK"
complement (6526..8850)

ORIGIN
Query Match 20.5%; Score 64; DB 1; Length 10029;
Best Local Similarity 50.7%; Pred. No. 5.1;
Matches 154; Conservative 0; Mismatches 150; Indels 0; Gaps 0;
/note="orf30"
/codon_start=1
/transl_table=11
/product="5-methyltetrahydropteroylirriglutamate-
homocysteine methyltransferase-like protein"
/protein_id="AA081245.1"
/db_xref="GI:8896117"
/translation="MTAKPAAAAARATVYGYPROGPNRLKRAIBGWKGRVDAALR
ETRAELRRGTWQOLAEGHNDPTEGPFSTYDHYLDTSVWYAGVPGHRAAVDADLDG
YPMAGCTDVAPELEMTKMPDNYHYLVEPELGPVFTDASAKOVAECALAGHTP
RPLVGPVYLLAKPAPVADFEPLTLDRLLPYAEVLADLRAAGAAVQDLDEPA
LVQDTPPAELNAAARAYRELIGLADRPOLIVASYFGLRGALTVLAKAPDGLAFT
ESGAGLDDLAAGGLPGKRLVGVVNGRNIWINDYKSLATGTLLGLADRVAS
CSLIHPLADADERDIPOLIAWLAFAARQSEIYVLAAGLTGGTDTAEIARNAD
LSRAAALITRDPARTATATADAGRSQPTTESAORAHGLPILPTTIGSP
QTTELRTARADLAKRIDAGYERIKDEIRVLSYQEARGLVHVEPERNDVOY
FAPOLTYLATOHGWQVOSYGRVYRPVYLAGDISREPMVWTTVAQSLTAPVKM
LNGPVTLAMS FVRDQPLGDPARQVALARBEVNDLEASGTVIOVDPALRETLPL
RAEHAAYLEMATSEFRLATGVRPDTOIHTHWCYAEFGDIYQALIDLDADVLSLEA
RSHMOYARELAAGYPERAGPGVYDIHSRPISTEEAALLRKGLAIPASRLMVPND
GLKXIGWETTRASLENVAAAREIARLPTAEA"
RESULT 11
AE005068/c 13635 bp DNA linear BCT 12-FEB-2001
DEFINITION Halobacterium sp. NRC-1 section 99 of 170 of the complete genome.
VERSION AE005068.1 GI:10581033
KEYWORDS
SOURCE
ORGANISM Halobacterium sp. NRC-1
Halobacterium sp. NRC-1
Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Halobacterium.
REFERENCE
AUTHORS Ng,M.V., Kennedy,S.P., Mahaffas,G.G., Bergquist,B., Pan,M.,
Shukla,H.D., Iasky,S.R., Baliga,N., Thorson,V., Strogon,J.,
Swartzell,S., Weir,D., Hall,J., Dahl,T.A., Weitz,R., Geo,Y.A.,
Leithauser,B., Keller,K., Cruz,R., Danson,M.J., Hough,D.W.,
Maddocks,D.G., Jablonksi,P.E., Krebs,M.P., Angevine,C.M., Dale,H.,
Isenbarger,T.A., Beck,R.F., Pohlischrod,M., Spudich,J.L.,
Jung,K.-H., Alam,M., Freitas,T., Hou,S., Daniels,C.J., Dennis,P.P.,
Omer,A.D., Ehardt,H., Lowe,T.M., Liang,P., Riley,M., Hood,L. and
DasSarma,S.

TITLE Genome sequence of Halobacterium species NRC-1
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)
MEDLINE 20504483
PubMed 11016950
REFERENCE 2 (bases 1 to 13635)
AUTHORS Ng, W. V., Kennedy, S. P., Mahajras, G. G., Berquist, B., Pan, M., Shukla, H. D., Laskey, S. R., Baliga, N., Thoreson, V., Sbrogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T. A., Weir, R., Goo, Y. A., Leithauser, B., Keller, K., Cruz, R., Danson, M. J., Hough, D. W., Maddocks, D. G., Jablonski, P. E., Krebs, M. P., Angevine, C. M., Dale, H., Isenbarger, T. A., Peck, R. F., Polischod, M., Spudich, J. L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C. J., Dennis, P. P., Omer, A. D., Ehardt, H., Lowe, T. W., Liang, P., Riley, M., Hood, L. and Dassarma, S.
TITLE Direct Submision
JOURNAL Submitted (14-JUL-2000) Institute for Systems Biology, 4225 Roosevelt Way NE, Seattle, WA 98105, USA
FEATURES
Source Location/Qualifiers
1.13635
/organism="Halobacterium sp. NRC-1"
/mol_type="genomic DNA"
/strain="NRC-1"
/db_xref="taxon:64091"
complement(142..633)
/gene="VNG1547C"
complement(142..633)
/gene="VNG1547C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1547c"
/protein_id="AAG19830.1"
/db_xref="GI:10581034"
/translation="MDRVSKASRIEAYGVDELNAVGAIRPTDIADLWLREXON
HHVVOADPANPEPDPDPOITAEHTYKEMIDHCERLDPLESTILLGGSEAGAL
HARAVACRAERAAVALQDHEPDANDAVAVLNRLDALTVFARLADVAGVHES
PST"
complement(681..920)
/gene="VNG1546H"
complement(681..920)
/gene="VNG1546H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1546h"
/protein_id="AAG19831.1"
/db_xref="GI:10581035"
/translation="MTLTLYALDGGCPDCESVITDLAADDIDHETVHVDAHRSARNAVK
RASGQRSVVYLDDBRGVVMADSORIQTTAAVTLA"
complement(1776..2588)
/gene="VNG1548C"
complement(1776..2588)
/gene="VNG1548C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1548c"
/protein_id="AAG19832.1"
/db_xref="GI:10581036"
/translation="WYESRANGDLGTLRKRNATRYQIIIVETAEROPVAVSOEIADEI
GYTAQVNSDYLDLVEEGHARKRGKRYVTEGVNWLISQTDLLTFVEYVESVLE
EYVDVTAADADIDIGERYLSMHDGVMHATPGADGSATAVAVGAAACDGTAEFD
ENLDYEPGEVTVLSTFVVODGSGALPPDDAVTARLAADHIVAAGVAAVAVRDA
GSEPIRRTGASAVPAASRGSLVLLVTADRLSHTDPLRERNLYGEVVGGLD"
2727..3290
/gene="cbiI"
/note="synonym: VNG1550G"
2727..3290
/gene="cbiI"
/note="cbiI"
/codon_start=1
/transl_table=11
/product="cobalamin biosynthesis"

gene
CDS
/protein_id="AAG19833.1"
/db_xref="GI:10581037"
/translation="MAQQLPHDAKAGPTKPEVRAVVRSKLATLTDHFEVSGCTGA
VTIAAADRGVTLAEKADRLRETKRNLAAAGLADAVELRAVEAPAEELPADADLFI
GSRVYDALDHAETGVDRIYMNYSRLVEVNAAGLQAQPRDRDILTEVVGFOYSHGTEL
AAATSPDSNPFYMLVGSASGTEADR"
3287..3991
/gene="cbiI"
/note="synonym: VNG1551G"
3287..3991
/gene="cbiI"
/note="cbiI"
/codon_start=1
/transl_table=11
/product="cobalamin biosynthesis"
/protein_id="AAG19834.1"
/db_xref="GI:10581038"
/translation="MTVVGIGLGPREDLLITRGKRTLEADVYVSPGRLSSTVATTH
VPDSRIQDIDFPMTSDPEKRLAMRDAAAEVAPRARDDVAFATIGDPVYVSTFGLR
RTLDAFHPDIDQTVGVSAVSAFTTALGVEVDAGAGALREAAQASPTGDMILF
KVTDAPTHGGLDAGDVTVYGRRLYMEQDVTYTRDPAELADRDYTLAYMRDTLQ
KDRATAPDTEVAGDA"
3984..4850
/gene="cbiF"
/note="synonym: VNG1553G"
3984..4850
/gene="cbiF"
/note="cbiF"
/codon_start=1
/transl_table=11
/product="cobalamin biosynthesis"
/protein_id="AAG19835.1"
/db_xref="GI:10581039"
/translation="MPEDATPDQADDAASQORGRDARVYDHTAGDPOAGIPIFAG
PQDRLTLVYKSGVEDADIVYHAGSLVNSLDDYKDAEYVSGVDLIELTLMR
DAHEGDTVVRHSGDPAIYGAALEOMALBQSGPTTYVGEVTSFASATLRTQLT
LNGIANHVAFTRPQGTLDABEDDIAAEFVEGMDVTTCYLGTHALPYMERLLAAGHD
PBTPTVYVYHASMWPDEDVIEGTIETIGDREDAGYRASALAVTGDAGAGYERSYLY
GDMANRSGSDTDQ"
4847..5824
/gene="cbiG"
/note="synonym: VNG1554G"
4847..5824
/gene="cbiG"
/note="cbiG"
/codon_start=1
/transl_table=11
/product="cobalamin biosynthesis"
/protein_id="AAG19836.1"
/db_xref="GI:10581040"
/translation="MSTNDTETDSDGSGHSTPDSGDGEVAEDIAIVAFERRKRTAAEI
VAGIGDRYDSIDILDHYGDVFAEHGVDGCVGLMASGIAMRKTGILLDDKEDPAVV
VYDEALTWAITPLITGHRGANOVADDLASMGAVPANTTSAAGKGVSEKAKALDTHV
VNGDSTVATNLVLDENLGPVAKLDGPVAVVGDVTVLKNSDGGVVLGTSVSGAD
EQQFPRDAAASLAACTDYDIDVEFVATRKRADEGLAAAEALGLGVVAPEKTLLEA
HEGPTSPRSRKEILGWPGIAEASAIAGRDNELLAEKERYDEAVTAAVGK"
5821..6651
/gene="cobH"
/note="synonym: VNG1555G"
5821..6651
/gene="cobH"
/note="cobH"
/codon_start=1
/transl_table=11
/product="cobalamin biosynthesis"
/protein_id="AAG19837.1"
/db_xref="GI:10581041"
/translation="MSDANEADYGTLLVVGIGPLPGAMTORAKOVVVTADCVVANSN
LYOERLRDGTLLPPADAGERPDQEVYRS TMGRQIELAEFAERYVDSGDVAHVSGDP
NVYGSQSDVYLAABEBAAYDVPIELVPGVTALGAALUGAPLSNDFCTISLSDKRG
WEBIEKRAAATISGFVVLVYCMWRDVRATGALREERADVPVATINDAGDGGGRN
TDGETHTITTLGSVTDSHGVEGMGASLITGIDETAVMENDHRRKLLVTPRGGRDVEDF"

gene	6651..7658	/gene="cbiH"	/note="synonym: VNG15576"
CDS	6651..7658	/gene="cbiH"	/note="cbiH"
		/codon_start=1	/transl_table=1
		/product="cobalamin biosynthesis"	/protein_id="AAG19838.1"
		/db_xref="GI:10581042"	/translation="MSITDITTDSTNSCGASTTDDSTSSCGGATTDDBASSCGALAAAEHIVGTYTVDLLPDEITDDAEIYDTPMCGBSRTEBYADRLAENHYAIIIGSGDNNVVALAGLALEITESKATPSMDFEVPGVPAAGSCGARLGAFLVNDVSISLDHILTPMEIESRLHVAVEEFTIISVNPMSRKRENFQKCEIILAHRDADTPVGIVHGAEREDEVEIITLGELESYGETDLIDMTTVIYIGNAFYFDDRMVTPRGYETKYDY"
gene	7658..7906	/gene="VNG1558H"	
CDS	7658..7906	/gene="VNG1558H"	/note="hypothetical protein"
		/codon_start=1	/transl_table=1
		/product="Vng1558h"	/protein_id="AAG19839.1"
		/db_xref="GI:10581043"	/translation="MYRTVIDREACDGVFACLTDRSFEVDAGLATLADDDGVERTRKETVYGFDDDERVADADDAACFPDAISDAVADE"
gene	7903..8601	/gene="VNG1559H"	
CDS	7903..8601	/gene="VNG1559H"	/note="hypothetical protein"
Query Match	19.5%;	Score 60.8;	DB 1; Length 13635;
Matches 140;	Conservative 51.5%;	Prod. No. 13;	Mismatches 132; Indels 0; Gaps 0;
Oy	35	CGGAGCTCGAGCCCTCGAGGGAAGCTCCCTCGACACGGTGGCGCGCATCGACCTCG	94
Db	545	CGGACATGCCGACCTCGACGCTTGCTCCGGAGGCAACAGACCACTTCACGCTGTC	486
Oy	95	GTCGCCGATCGCGCGGTACAAGTCCCGGACCGCGCTCCGATGATGACGCGCGG	154
Db	485	AGGGGAACTTGGCAACCCCGAGCCCGAGCCGACCAAGATCACCGCCGAACAG	426
Oy	155	TCAGCTTGTCAGACAGGCGCGCGCTACGCGCCGACCAAGCGCTTCAGCAATGCT	214
Db	425	TCACCAAGTCGAGAGTGGATTCGACCACTGCGAGGCGCGCTGGACCCCTGAGTGT	366
Oy	215	TCCCTGTAACCTCTACGACGTGATCATCAGGAGATGTGCGCGCTCGAGGACCTGGTGA	274
Db	365	TCATCTGCGCCGGGGGCGACGAGGCGGCGCTGTCATCAGCGCGCGGTGTGCC	306
Oy	275	TGAGCCGGAGAGCCTGACGCGCCGAGACCGG	306
Db	305	GGCGCGCGGAGCGCGCGCGGTGCGGCTCCAG	274
RESULT 12	AK121213	1964 bp	mRNA linear
LOCUS	AK121213		PLN 29-OCT-2003
DEFINITION	Oryza sativa (japonica cultivar-group)	cdna clone:J023089K20, full	
ACCESSION	AK121213		
VERSION	AK121213.1	GI:37990836	
SOURCE	Oryza sativa (japonica cultivar-group)		
ORGANISM	Oryza sativa (japonica cultivar-group)		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;		

REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED	REFERENCE AUTHORS	TITLE	JOURNAL MEDLINE PUBMED			
1	Emhartoidaeae; Oryzeae; Oryza.		2	Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Doi, K., Fujimura, T., Fukuda, S., Hangan, T., Hara, A., Hashizume, M., Hayashida, K., Hayashizaki, Y., Hayatsu, N., Hiramoto, K., Hirooka, T., Hori, F., Hotta, I., Iida, J., Iida, Y., Ikeda, R., Imamura, K., Imotani, K., Ishibiki, J., Ishii, Y., Ishikawa, M., Itoh, M., Kagawa, I., Kawanaga, S., Katoh, H., Kawagashira, N., Kawai, J., Kawamata, M., Kikuchi, S., Kishikawa-Hirozane, T., Kishimoto, N., Kobayashi, M., Kodama, T., Kojima, K., Kojima, Y., Kondo, S., Konno, H., Kouda, M., Koya, S., Kurihara, C., Kurosaki, T., Kusumegi, T., Li, C., Lu, M., Masuda, H., Matsubara, K., Matsumura, T., Miura, J., Miyazaki, A., Mizuno, K., Murakami, K., Murata, M., Nagata, T., Nakahama, Y., Nakamura, M., Nani, T., Narikawa, R., Nikura, J., Nishii, K., Nomura, K., Numasaka, R., Ohneda, E., Ohno, M., Ohtsuki, K., Oka, M., Ooka, H., Oosato, N., Ota, Y., Otomo, Y., Ryu, R., Satoh, H., Sakai, C., Sakai, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Satoh, K., Shibata, K., Shinagawa, A., Shiraki, T., Shishiki, T., Sogabe, Y., Sugano, S., Sugiyama, A., Suzuki, K., Suzuki, Y., Tagami, M., Tagami, Takeda, Y., Tagawa, A., Takahashi, F., Takaki-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Tsunoda, Y., Ueda, M., Waki, K., Xie, Q., Yahagi, W., Yamada, H., Yamamoto, M., Yasunishi, A., Yazaki, J., Yokomizo, S. and Yoshimura, A.	Science 301 (5631), 376-379 (2003)	1	The Rice Full-length cDNA Consortium, National Institute of Agricultural Sciences Rice Full-length cDNA Project Team, Kikuchi, S., Satoh, K., Nagata, T., Kawagashira, N., Doi, K., Kishimoto, N., Yazaki, J., Ishikawa, M., Yamada, H., Ooka, H., Hotta, I., Kojima, K., Nami, T., Ohneda, E., Yahagi, W., Suzuki, K., Li, C., Ohtsuki, K., Shishiki, T., Foundation of Advancement of International Science Genome Sequencing & Analysis Group, Otomo, Y., Murakami, K., Iida, Y., Sugano, S., Fujimura, T., Suzuki, Y., Tsunoda, Y., Kurosaki, T., Kodama, T., Masuda, H., Kobayashi, M., Xie, Q., Lu, M., Narikawa, R., Sugiyama, A., Mizuno, K., Yokomizo, S., Nikura, J., Ikeda, R., Ishibiki, J., Kawamata, M., Yoshimura, A., Miura, J., Kusumegi, T., Oka, M., Ryu, R., Ueda, M., Matsubara, K., Riken, Kawai, J., Carninci, P., Adachi, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, M., Hayatsu, N., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kondo, S., Konno, H., Miyazaki, A., Oosato, N., Ota, Y., Saito, R., Sasaki, D., Sato, K., Shibata, K., Shinagawa, A., Shiraki, T., Yoshino, M. and Hayashizaki, Y.	Science 301 (5631), 376-379 (2003)
2	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	3	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	3	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
3	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	4	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	4	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
4	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	5	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	5	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
5	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	6	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	6	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
6	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	7	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	7	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
7	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	8	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	8	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
8	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	9	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	9	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)
9	Collection, mapping, and annotation of over 28,000 cDNA clones from japonica rice	Science 301 (5631), 376-379 (2003)	10	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences, Department of Molecular Genetics, Head of Laboratory of Gene Expression; 2-1-2 Kamondai, Tsukuba, Ibaraki 305-8602, Japan (E-mail: kikuchi@nias.affrc.go.jp, Tel:81-29-838-7007, Fax:81-29-838-7007)	Science 301 (5631), 376-379 (2003)	10	Submitted (31-JAN-2003) Shoshi Kikuchi, National Institute of Agricultural Sciences,	

Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hanagaki, T., Hara, A., Hashizume, W., Hayashida, K., Hayatsu, N., Hiramoto, K., Hiroaka, T., Hori, F., Iida, J., Imamura, K., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kanagawa, S., Kato, H., Kawai, J., Kishikawa-Hirozane, T., Kojima, Y., Kondo, S., Konno, H., Konda, M., Koya, S., Kurihara, C., Matsuyama, T., Miyazaki, A., Murata, M., Nakamura, M., Nishi, K., Nomura, K., Numasaki, R., Ohno, M., Otsu, N., Ota, Y., Saitoh, H., Sakai, C., Sakai, K., Sakai, K., Shibata, K., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Sato, K., Shikama, A., Shiraki, T., Sogabe, Y., Tagami, M., Tagami-Takeda, Y., Tagawa, A., Takahashi, F., Takaku-Akahira, S., Tanaka, T., Tomaru, A., Toya, T., Waki, K., Yasunishi, A. and Hayashizaki, Y.

Location/Qualifiers

1. .1964

/organism="Oryza sativa (japonica cultivar-group)"

/mol_type="mRNA"

/cultiyar="Nipponbare"

/db_xref="taxon:39947"

/clone="U023089K20"

ORIGIN

Query Match 19.0%; Score 59.4; DB 8; Length 1964;

Best Local Similarity 50.2%; Pred. No. 32; Matches 147; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 16 GAGCTGACAGCGGCTGCGCGCGAGCTCGAGCCCTCGACGGAGCGCTCTGACACGGTG 75

Db 967 GAGGAGAGAGAGAGAGCGCCCTCGTCACTGCTGCTGGCGGGGTCGACCGGGGTC 1026

QY 76 CGCGCGCGCATCGACCTCGGTGTCGGATGCGCGGTACAGTCCCGGACCGCGTCCG 135

Db 1027 GAATTAGCGCGGACGCTCAGCGACTTCATCACCCTGGAGCTGGAGAGAGTACGCTCAC 1086

QY 136 ATGATGACAGCCCGCGCGGCTGCTGCTGACAGACAGCGCGCGCTACCGCGCGAC 195

Db 1087 GTCAAGAGACTACGCTCAAGCTCACCCTCATGAGAGCGACAGATCTCTCTCTTCGAC 1146

QY 196 CACGCGCTCGACGATGCTTCTGCTGTAACCTTACGACGCTGATCAACGAGATGTG 255

Db 1147 GTGCGGCTGCGCGGACGAGACGACGACCTCTTCAAGTAGCGCGTGAACCTGTGCGG 1206

QY 256 CGGCTGAGAGACTGCTGATGAGCCCGGAGAGGCTTACGCGCGAGACCGCGG 308

Db 1207 GCGGTGTGTAAGGAGGTGAAGCGAGGAGATGACGTGACGACGCGAGCGG 1259

RESULT 13

AB005050/c

LOCUS AB005050 16905 bp DNA linear BCT 12-FEB-2001

DEFINITION Halobacterium sp. NRC-1 section 81 of 170 of the complete genome.

ACCESSION AB005050 AB004437

VERSION AB005050.1 GI:10580768

KEYWORDS

SOURCE

ORGANISM Halobacterium sp. NRC-1

Halobacterium sp. NRC-1

Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Halobacterium.

1 (bases 1 to 16905)

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Bergquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N., Thorsen, V., Sirogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welt, R., Goo, Y.A., Leitbauer, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddock, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isehaber, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

TITLE

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (22), 12176-12181 (2000)

MEDLINE 20504483

PUBMED 11016950

REFERENCE 2 (bases 1 to 16905)

AUTHORS

Ng, W.V., Kennedy, S.P., Mahairas, G.G., Bergquist, B., Pan, M., Shukla, H.D., Lasky, S.R., Baliga, N., Thorsen, V., Sirogna, J., Swartzell, S., Weir, D., Hall, J., Dahl, T.A., Welt, R., Goo, Y.A., Leitbauer, B., Keller, K., Cruz, R., Danson, M.J., Hough, D.W., Maddock, D.G., Jablonski, P.E., Krebs, M.P., Angevine, C.M., Dale, H., Isehaber, T.A., Peck, R.F., Pohlschrod, M., Spudich, J.L., Jung, K.-H., Alam, M., Freitas, T., Hou, S., Daniels, C.J., Dennis, P.P., Omer, A.D., Ebhardt, H., Lowe, T.M., Liang, P., Riley, M., Hood, L. and DasSarma, S.

TITLE

JOURNAL Submitted (14-JUN-2000) Institute for Systems Biology, 4225

Roosevelt Way NE, Seattle, WA 98105, USA

FEATURES

source

1. .16905

/organism="Halobacterium sp. NRC-1"

/mol_type="genomic DNA"

/strain="NRC-1"

/db_xref="taxon:64091"

125. .346

/gene="VNG1238C"

125. .346

/gene="VNG1238C"

/note="conserved hypothetical protein"

/codon_start=1

/transl_table=11

/product="Vng1238C"

/protein_id="AAG19601.1"

/db_xref="GI:10580769"

/translation="MCDVDANGLVAQLARSLDPTFLIHTTPSRAPAFERLADI

ITTPHALAAATVDSHPAPASTHNSENA"

333. .2519

/gene="yhdg"

/note="synonym: VNG1240G"

333. .2519

/gene="yhdg"

/note="yhdg"

/codon_start=1

/transl_table=11

/product="amino acid transporter"

/protein_id="AAG19602.1"

/db_xref="GI:10580770"

/translation="WKTLEKRDGLGAVPAISVGMITGSGIFLTPALAKMAPAVIA

YLLAGLVVPAALSSSEMTAMPENGTYIYIERMGPLIGVAGVGMFSLSFSAL

ALVGVPLVPLFLFDLPVQVALGLAALIVNLLGAKOTGLQTVIVLMGALGMFV

AGSAGVOOANOQGLTSGMEGLAATGIVFVSYVAGVKVSVAREVNPBNPLGI

IGSLAFTLLVYGIYVAVGVTDGSAVAGVSAATPVVADATLTGTAGVAAVTAAYAL

VSTANAGLSSRRYPFAMSRDQAPPTPAETHDRGPTTTATLTGAVVALLIAFPI

LDIAKLASAPQILVFLVNLAVARTGTDDPDESPLVPMQGFVLSGLALDTG

MGWVPIAGAAVITLGSVVMRYVYARPRVBRGAATDVVRRRGVLTETBALVDATG

AEAGPEVILVALTREASADREBALISLAADRAADGQVVVRFVDDQAPLDHAD

VQSPADVFEERTBGLSTEFVDVLAIGVSVSHDKRAILINVAEORSVDTVAEHERLR

LSRRLGDPIMVYVHAPCDVILVBNRGVDAPOVYVLESGDGSPNATVAAVADVAGA

AGCTVTLGPIUSADHPDKQETVHAAYQDELAGLSAPVRLTDLRADGSDQTPADLVLR

TGAHRRVGGKSGSHDTADCTELTVYPHESRPQLRLARLIVE"

2595. .3371

/gene="sure"

/note="synonym: VNG1241G"

2595. .3371

/gene="sure"

/note="sure"

/codon_start=1

/transl_table=11

/product="stationary phase survival protein"

/protein_id="AAG19603.1"

/db_xref="GI:10580771"

/translation="WDADPEPILVTNDGDIDAPGIRALADGLDAGVNTVVAADNQS

ATGRMSQEVAVHDHLEGAVEGTPADCEVGLALGPPLVVGSGVEGNGLGMVY

GRSGVSAVAAAFGVPAIAYSMYREGEQGPAAVADYHAADATTHLADAVTGD

IFPTADYLVNVAHPDADAGEMVYTRSHAYDMTAQTGTVLVLYRLMAMMAAGDI

HDPOGTDRAVLDDGVSVSPITAFRSTHHDALDGIATEP"

complement (3614. .3880)

/gene="VNG1244C"

complement (3614. .3880)

CDS

gene

CDS

```
/gene="VNG1244C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1244c"
/protein_id="AAG19604.1"
/db_xref="GI:10580772"
/translation="MTLDELRAEIESIDREIVDLIARTVVAETIADVKAEGSLSTD
ESQEDAMERAGERRAAAFQVDNIVKAVFRLLIETNKIEQRETR"
complement (3877..4731)
/gene="VNG1245C"
complement (3877..4731)
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1245c"
/protein_id="AAG19605.1"
/db_xref="GI:10580773"
/translation="MDGRAVAPAGTVLNLATGVGAAPALDIDTEASVSTPESGV
SGRIAGHPADTALVERCVRIIDRYGDCGGHRTSEVPLAAGLKSSAAANATVL
ATLDALGVADENVDRADARIGVQARADAGTVTGAFDPAASMLGVAITDNEDDL
FRDAVEHMAAVMTPPERAYADADVACERVSGLAEHVAALAAAGDYATMTVGLAF
CAALDPPTAPAVTALPHAAGVSLSGTSPYVAVGDEBDGIEVSTRMHENGTVARETT
QLAGARTT"
4819..5247
/gene="VNG1246H"
4819..5247
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1246h"
/protein_id="AAG19606.1"
/db_xref="GI:10580774"
/translation="MSARNDIAPSTLCEVLAEEGVVEYTDGREAIFYNGVEPIETGL
HTNKGKDVHLVNTDPTETEGVLYVNDLKTGHDELSTEGGRILLDGERBEELPPGVT
VQADGVAILDIDDEPTARGVFVFAEDBELSERVAIEYIPRA"
complement (5405..6064)
/gene="Ybhf"
/note="synonym: VNG1247G"
complement (5405..6064)
/gene="Ybhf"
/note="Ybhf"
/codon_start=1
/transl_table=11
/product="ABC-type transport protein"
/protein_id="AAG19607.1"
/db_xref="GI:10580775"
/translation="MITATDITKTYGETTALAASVAMDSGLHCLAPNDSKTTLLR
VLGLTRHDSGSVTPESLGVAFQTPNVYADLVTAENLDVFGALAGATDWRASLVQ
LRLDVPAGRTAGASGGFQKLDLALALRPAALLDEPLADDPATRRRLVDTVTR
YSAHTVVASTHLDAPADYATLTVVRDGLLDSMDRPAAPNDPSGVYRLDRENG
F"
complement (6067..7137)
/gene="VNG1249C"
complement (6067..7137)
/gene="VNG1249C"
/note="conserved hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1249c"
/protein_id="AAG19608.1"
/db_xref="GI:10580776"
/translation="MDVSLKGLRWGRHKLALAAVFLVPAVAVTVGFOHLP
TDIAVAPRTAAVTGDLRLYRAAAVTFSDPVYGRADLTLTGEOYVAAVSVRP
ALADSARGAAFTYIDGSVVPHEPSRAIYRLIQSLGTAIPDYTVSRFVCTOYS
LSTLVSPFVSLTAVLGYLPOTFEBAERDLKRETSLSIAVVASKLFPRALLV
VPLVFPADVAVTAGSVSLGAVATLAVVVMFLGLVAVALIQLAIFGWRGLTLNL
VLFLVLSGVSVPVGFSPARRWVRQVPTHYAVVTIRGLVLRGSPISAYADWLGCL
VAAVGAALVAGAIIELYERRA"
complement (7138..8229)
/gene="VNG1250H"
complement (7138..8229)
```

```
CDS
complement (7138..8229)
/gene="VNG1250H"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="Vng1250h"
/protein_id="AAG19609.1"
/db_xref="GI:10580777"
/translation="MYRDLHVPRTALLAVGIAVLVAAQGTGMRGSSADATTNKL
DSVPASAEYAVYDVAGMTDIDLRSVTTTTLAEVVLGSRAPSDVSSKMLDFFDADI
NPKRIQNTMTFTASTIATQPTGSSGAIVLSTYDESGVNALESSTDTLLNEGYGNV
TVYSDSLGFDVNLALADGERFVWDKRSVESMLDVRSGDADHLGDSAFEGVNG
GSYYQALSAFNGSDSLIGYERVAENAPVNLASVSGIIEYTAVSFGTDDGDITSEHY
ADESGAEARAYDIDGTRLYASPEEELRPSPEAGSIEQDQNTVIVTVAESADTIES
RIEARGTNETADNTTLLDNSSAASVARE"
complement (8313..11153)
/gene="hela"
/note="synonym: VNG1251G"
complement (8313..11153)
/gene="hela"
/note="hela"
/codon_start=1
/transl_table=11
/product="ATP-dependent helicase"
/protein_id="AAG19610.1"
/db_xref="GI:10580778"
/translation="MSETGGRWDAFTHLGSEVRNALSRRGFTTPTPEPORAVPALD
GRNLVAPFTGTGTETAMLVYDALVSDGFPDGCALVTPTLRALNDRMRRLDM
GETLGLAEVDVRHGDITTDYQRQDPTDPDLVTPPTLQAMLGSLRLGLADVEHY
VDEHBLAGAKRGAQLAIGLERLVEVAGVQRIGLSATYQDPHEIGAFLTGRGCEII
```

Query Match 18.7%; Score 58.2; DB 1; Length 16905;
Best Local Similarity 53.8%; Pred. No. 27; Matches 120; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

Qy	28	CTCGCGCGGAGCTTCGACGCGCTCTGACGCGGACCTCTCGACACGGTGC	87
Db	3865	CTCGCGCGGAGATTCGACGACATCGACCGGAAATCGTGAATCGCGGACGACG	3806
Qy	88	GACCTCGGTGTCGCGATCGCGCGGTACAGTCCCGGACGCGCTCCGATGATGACGCC	147
Db	3805	TACGTGCGCGAGACCAATCGCGCAGTGAAGCCGACGAGGCGGTGTCGACGACGACGAG	3746
Qy	148	GCGCGGTGACCTGTGTCAGACGACGAGCGCCCGCTACGCGCCGACCAAGCGCTCGAC	207
Db	3745	TCCGAGGAGGACCGACGATGATGAGCGCGCCGCGAGAACCGCGGCTTCGCGGTGAC	3686
Qy	208	GAATCGTTCCTGGTGAACCTCTACGACGTGATCATCAGGAGA	250
Db	3685	GACCACTCGTGAAGCGGTGTTCCGGCTCTCATCGAGTTGA	3643

RESULT 14
AP006840_33/c
WPCOMMENT

Sequence split into 36 fragments	LOCUS	AP006840	Accession	AP006840
Fragment Name	Begin	End		
AP006840_00	1	110000		
AP006840_01	100001	210000		
AP006840_02	200001	310000		
AP006840_03	300001	410000		
AP006840_04	400001	510000		
AP006840_05	500001	610000		
AP006840_06	600001	710000		
AP006840_07	700001	810000		
AP006840_08	800001	910000		
AP006840_09	900001	1010000		
AP006840_10	1000001	1110000		
AP006840_11	1100001	1210000		
AP006840_12	1200001	1310000		
AP006840_13	1300001	1410000		
AP006840_14	1400001	1510000		
AP006840_15	1500001	1610000		

Query Match	17.9%;	Score 55.8;	DB 1;	Length 110000;
Best Local Similarity	55.4%;	Pred. No. 36;		
Matches 108; Conservative	0;	Mismatches 87;	Indels 0;	Gaps 0;

b 29025 GAGGTCGGGAGATC 29011

AL939120.1 GI:24413871

Streptomycineae; Strept

AUTHORS Bentley, S.D., Chater, K.

Parkhill, J. and Hopwood

JOURNAL Nature 417 (6885), 141-

PUBMED	12000953
--------	----------

TITLE	Direct Submission

Source

gene

CDS

gene

CDS

misc_feature

genes

```

CDS
/note="synonym: SC6F11.12"
/2319..3644
/gene="SCO4414"
/note="SC6F11.12, possible integral membrane protein, len:
441 aa; similar to TR:BA06413 (EMBL:AP001516) Bacillus
halodurans transporter BH2694, 418 aa; fasta scores: opt:
795 z-score: 909.8 E(): 0; 31.0% identity in 413 aa
overlap. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane transport protein"
/protein_id="CAC08423.1"
/db_xref="GI:10178385"
/db_xref="UniProt/TREMBL:Q9F374"
/translation="MTOTDAAAAAASERPEPTRGPTRAHRAWVAATPVITII
GAAPRSLPGILIDPLHOFGRSGRTGAASVVALYGLTAPPAALMDRFGIRRV
AVALLMVALSGSLTVMNNSAQILICGLVGLSGMALAFATVTNRWTFRRGLV
TGVLTASASGQILFLPLSMIVETVDMRPAAYVALAALAVPFWLLDHPADV
QKPYGAKEFVKKPAPVPGARBALTVFAVRGSPFWLLAGTALICGASTNGLIOTHF
VPAADHGMRTTAAASLAIYGVFDVAGTASGFTDRFEPRLIAYVALRGVSLF
LPMULAPVHPMIFIVFYGLDVAIVPPTLALCRHGEDSAIVFGVWLASHQVGA
ALVAFLEGVARDFGSYDVVVIASGALCAAAALMALVIRRRPVVALS"
/3666..4538
/gene="SCO4415"
/note="synonym: SC6F11.13"
/3666..4538
/gene="SCO4415"
/note="SC6F11.13, possible integral membrane protein, len:
290 aa. Contains possible hydrophobic membrane spanning
regions"
/codon_start=1
/transl_table=11
/product="putative integral membrane protein"
/protein_id="CAC08424.1"
/db_xref="GI:10178386"
/db_xref="UniProt/TREMBL:Q9F373"
/translation="MGCTPEPAEAGRPPLMSDENVVAALCATQPLMLAWRFV
GAQDDHGRGGAFFVLCPLVLPGLSVASSAQIMPAATLALPLRRSGPEARHV
ASVVGAGMAVGLHWMPRPDTPVPAAGAILPVFLPLRRRMGMSWLRASG
ACPLVASGGLAVALADEYEPALSSGQALGEMRGHGVLLAPGCRGLRVSF
PDCGTADYTRCGGTGWSRNLGDRGGVVRLLDQCGEETSWTIGSSERSPLFALF
GDDADGDLFLTN"
/complement(4721..5926)
/gene="SCO4416"
/note="synonym: SC6F11.14c"
/complement(4721..5926)
/gene="SCO4416"
/note="SC6F11.14c, probable monooxygenase, len: 401 aa;
similar to TR:Q9SD6 (EMBL:AP000837) Oryza sativa similar
to human dimethylallylne monooxygenase, 437 aa; fasta
scores: opt: 692 z-score: 722.8 E(): 9.3e-33; 38.2%
identity in 403 aa overlap. Contains Pfam match to entry
PF00743 FMO-like, Flavin-binding monooxygenase-like"
/codon_start=1
/transl_table=11
/product="putative monooxygenase"
/protein_id="CAC08425.1"
/db_xref="GI:10178387"
/db_xref="GOA:Q9F372"
/db_xref="UniProt/TREMBL:Q9F372"
/translation="MADSTASAPQRPPTHAPRPVYVIGAPGGLAVAAHARGLRA
VVLERADHVSSWRHRYDLRLHTTRLSLPGILIPRRGRVAVADDVRYLEKTAE
YHOLEIVTGVEFRVERAPDGTGMLLHAAGRELGAIVVATGVNHTPPVPMGRD
TYTGEPHADYRTPAVAGDVLVGVGNTGABIAVDIVEGGAARVRLAVRTAPHI
RSTAGAAQYTGVLGRULPVALLVPLARIASIVDLSAQGLPRPGTGLSRVAG
AIRVQVGLIDAVRSGKEVVAAMDGFEDGKVLADGTRTAPDAVTAATGTRKGLSL
VGLGVLDGTGRPVVGGRTPAAPGLYFTGTFNPLSGMLRELAIDAERTAGAVARG
AGGVSRLLPG"
/complement(4754..5875)
/gene="SCO4416"
/note="Pfam match to entry PF00743 FMO-like,
flavin-binding monooxygenase-like, score -255.00, E-value
```

```

gene
8..8e-09"
6026..6643
/gene="SCO4417"
/note="synonym: SC6F11.15"
/6026..6643
/gene="SCO4417"
/note="SC6F11.15, probable acetyltransferase, len: 205 aa;
similar to SW:RIWL_ECOLI (EMBL:X15860) Escherichia coli
ribosomal-protein-terine acetyltransferase (EC 2.3.1.-)
RIWL, 179 aa; fasta scores: opt: 199 z-score: 261.1 E():
4.8e-07; 30.6% identity in 157 aa overlap. Contains Pfam
match to entry PF00583 Acetyltransf, Acetyltransferase
(GNAT) family"
/codon_start=1
/transl_table=11
/product="putative acetyltransferase"
/protein_id="CAC08426.1"
/db_xref="GI:10178388"
/db_xref="GOA:Q9F371"
/db_xref="UniProt/TREMBL:Q9F371"
/translation="MAHTPMYTLDESLESTERPRGHLRLPAMDAGSDDAHTVRLCL
DPEPQWNTTLKLTLDLAGARSLRGRADADGRASFRITDESGTTLGHLGMEI
SRQMKVARGVWVLPARGRGVATRALVASSWAVYELGHLRLGHAAGHDSCRVA
ERSGVVEGTLREAMEAGRHDAFRVHLHARLATDEBPTPGAR"
/6290..6523
/gene="SCO4417"

misc_feature
Query Match 17.9%; Score 55.8; DB 1; Length 24850;
Best Local Similarity 49.8%; Pred. No. 30;
Matches 141; Conservative 0; Mismatches 142; Indels 0; Gaps 0;

QY 9 GCAGAACGAGCTGCACGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGCTCCTGGA 68
Db 185684 GCTGTTCTGGCCCTCTCAAGCCCGCGAGGAGATCCTCGCCACCGGATGGT 185625

QY 69 CACGATGCGGCGCGGATGACCTCGATGCGGATGCGGATGCGGATGCGGATGCGG 128
Db 185624 GTGATCCTCGCTTGTGCGCGCTCGCTTGTGCGCGCTCGCTTGTGCGCGCT 185565

QY 129 CGTCCGATGATGACGCGCGCGCGGCTGACGCTGATCAAGACAGGCGCGCGCTACGC 188
Db 185564 CGAGCTGCTGCGGCGGACCGCGGAGGCTCGCGCTGCGGCGCGCGCTGCTAC 185505

QY 189 CGCCGACACGCGCTGACGATGCTTCTGTTGAACTCTACAGCTGATCAACGGA 248
Db 185504 CGTCACTGGGCGCTACATCGTGGCGCTGAAACAGCGCGCATGCTGAGGCGCT 185445

QY 249 GATGTCCCGCTGAGAGACTGTGATGACGCGGAGACCTG 291
Db 185444 CGGCTACTTCAACACCGCTGCTACATCGGATGGCGTG 185402

Search completed: October 5, 2005, 05:30:50
Job time : 1437.22 secs
```

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 16:46:09 ; Search time 188.488 Seconds
(without alignments)
9798.799 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312
Sequence: 1 atgcgcgagcagcagcagct.....cggcgcgagcagcgcggtga 312

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	312	100.0	312	4	AAf86410
2	312	100.0	312	8	AAf86410 4-amino-4
3	296.4	100.0	3305	8	AAf86410 4-amino-4
4	264.8	84.9	12391	8	AAf86410 4-amino-4
5	135.6	43.5	387	2	AAf86410 4-amino-4
6	135.6	43.5	2888	2	AAf86410 4-amino-4
7	54.8	17.6	1458	10	AAf86410 4-amino-4
8	54.8	17.6	45055	10	AAf86410 4-amino-4
9	54.6	17.5	507	11	AAf86410 4-amino-4
10	54.6	17.5	1359	11	AAf86410 4-amino-4
11	54.6	17.5	1371	11	AAf86410 4-amino-4
12	54.6	17.5	73882	13	AAf86410 4-amino-4
13	54.6	17.3	1208	2	AAf86410 4-amino-4
14	54.6	17.3	5392	2	AAf86410 4-amino-4
15	53.6	17.2	837	10	AAf86410 4-amino-4
16	53.6	17.2	1933	2	AAf86410 4-amino-4
17	53.6	17.2	29870	8	AAf86410 4-amino-4
18	53.6	17.2	45055	10	AAf86410 4-amino-4
19	53.6	17.0	1356	8	AAf86410 4-amino-4
20	52.8	16.9	510	11	AAf86410 4-amino-4

21	52.8	16.9	1116	11	ABD09634	ABD09634 Pseudomon
22	52.8	16.9	1125	11	ABD09470	ABD09470 Pseudomon
23	52.6	16.9	1869	13	ABD5461	ABD5461 Bacterial
24	52.6	16.9	1869	13	ABD54946	ABD54946 Bacterial
25	52.6	16.9	3735	13	ADT45696	ADT45696 Bacterial
26	52.4	16.8	1410	12	ADM80073	ADM80073 Spiroamyl
27	52.4	16.8	1410	12	ADN97589	ADN97589 S ambofac
28	52.4	16.8	30943	12	ADM80034	ADM80034 Spiroamyl
29	52.4	16.8	30943	12	ADN97550	ADN97550 S ambofac
30	51.8	16.6	1686	8	ADN6877	ADN6877 S. clavul
31	51.6	16.5	13654	10	ADC86678	ADC86678 Human GPC
32	51.4	16.5	388	9	ACL18903	ACL18903 DNA clone
33	51.2	16.4	313	4	AAC89250	AAC89250 Human bra
34	51.2	16.4	5469	2	AAV83491	AAV83491 Human T-t
35	51.2	16.4	5562	2	AAV83491	AAV83491 Human cal
36	51.2	16.4	5562	2	AAV83491	AAV83491 Human cal
37	51.2	16.4	6816	6	AAf31674	AAf31674 Human alp
38	51.2	16.4	6816	6	AAf31674	AAf31674 Human alp
39	51.2	16.4	6816	12	ABX93560	ABX93560 Human CDN
40	51.2	16.4	6816	12	ADH69264	ADH69264 Human TCC
41	51.2	16.4	6855	6	ABX3561	ABX3561 Human T-t
42	51.2	16.4	6855	12	ADH69266	ADH69266 Human TCC
43	51.2	16.4	6855	12	ADH69266	ADH69266 Human TCC
44	51.2	16.4	25467	12	ADQ07566	ADQ07566 Human vol
45	51.2	16.3	113193	8	AAf54645	AAf54645 Streptomy

ALIGNMENTS

RESULT 1	
AAf86410	
ID	AAf86410 standard; DNA; 312 BP.
XX	
AC	AAf86410;
DT	22-JUN-2001 (first entry)
XX	
DE	4-amino-4-deoxychorismic acid mutase coding sequence.
XX	
KW	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;
XX	4-amino-4-deoxychorismic acid mutase; enzyme; papB; ds.
OS	Streptomyces venezuelae.
XX	
PH	Key
FT	CDS
FT	Location/Qualifiers
FT	1..312
FT	/*tag= a
FT	/product= "4-amino-4-deoxychorismic acid mutase"
XX	
FN	WO200123542-A1.
XX	
PD	05-APR-2001.
XX	
PF	29-SEP-2000; 2000WO-JP06783.
XX	
PR	29-SEP-1999; 99JP-00276314.
XX	
PA	(MEIU) MEIJI SEIKA KAISHA LTD.
XX	
PI	Yana K, Okakura K, Yaeuda S, Watanabe M, Miyamoto K, Micho N;
PI	Murakami T;
DR	WPI; 2001-290517/30.
DR	P-PSDB; AAB82072.
XX	
PT	Transformant producing secondary metabolite modified with functional
PT	group e.g. benzene with nitrogen-containing substituent at para-position,
PT	PT1022, with ease at low cost, for application in pharmaceuticals and
XX	agrochemicals.
XX	
XX	Claim 13; Page 65; 83pp; Japanese.

Query	Match	Score	DB	Length	Matches	Conservative	Mismatches	Indels	Gaps
Qy	1 ATGACCGAGCAGACAGAGCTGCAGCGGCTGCGCGGAGAGCTGCAGCGCCCTGCAGCGGAGC	100.0%	DB 4	312	312	100.0%	0	0	0
Db	1 ATGACCGAGCAGACAGAGCTGCAGCGGCTGCGCGGAGAGCTGCAGCGCCCTGCAGCGGAGC								
Qy	61 CTCTGTGACACAGTGGTGGCGGCGCGCATCGACTCTGGTGTCCGATCGCGCGGTACAGTCC								
Db	61 CTCTGTGACACAGTGGTGGCGGCGCGCATCGACTCTGGTGTCCGATCGCGCGGTACAGTCC								
Qy	121 CGGCACGGCGCTCCCGATGATGACAGCCCGGCGCGGCTGCAGCTGTGCAGAGACAGGGCCGCC								
Db	121 CGGCACGGCGCTCCCGATGATGACAGCCCGGCGCGGCTGCAGCTGTGCAGAGACAGGGCCGCC								
Qy	181 CGGTACGCGCGCGCAGCAGCGGCTCGACGCAATGCTTCTGGTGAACCTCTACGACGTGATC								
Db	181 CGGTACGCGCGCGCAGCAGCGGCTCTCGACGCAATGCTTCTGGTGAACCTCTACGAGTGTATC								
Qy	241 ATCAGCAGAGATGTGTCGCGCTGCAGAGACCTGTGTATGAGCCGCGAGAGCGCTGACCGCCGAG								
Db	241 ATCAGCAGAGATGTGTCGCGCTGCAGAGACCTGTGTATGAGCCGCGAGAGCGCTGACCGCCGAG								
Qy	301 GACCGCGCGGTGA 312								
Db	301 GACCGCGCGGTGA 312								

RESULT 2

AA150181

ID AA150181 standard; DNA; 312 BP.

XX AA150181;

XX

DT 28-JAN-2003 (first entry)

XX

DE S venezuelae PF1022 substance gene #2.

XX

KW Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic; phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;

KW veterinary drug; gene; ds.

XX

OS Streptomyces venezuelae.

XX

Key Location/Qualifiers

FT CDS 1..312

FT FT /tag= a

XX /product= "PF1022 substance"

PN MO200277244-A1.

XX

03-OCT-2002.

XX

22-MAR-2002; 2002MO-JP002782.

PF

[illegible]

XX MO200285923-A2.
 PN 31-OCT-2002.
 PD 19-APR-2002; 2002MO-US012465.
 XX 19-APR-2001; 2001US-0285030P.
 PR 06-FEB-2002; 2002US-0355514P.
 XX (SCRI) SCRIPPS RES INST.
 PA Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
 PI Meggers EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;
 XX WPI; 2003-120430/11.
 DR
 XX Composition useful for producing protein comprising unnatural amino acid,
 PT has translation system comprising orthogonal tRNA and orthogonal
 PT aminoacyl tRNA synthetase.
 XX Example 4; Page 127-128; 188pp; English.
 PS
 XX The invention relates to a novel composition comprising a translation
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
 CC -tRNA with at least one unnatural amino acid in the translation system of
 CC and the O-tRNA recognises at least one selector codon. A composition of
 CC the invention is useful for producing at least one protein comprising at
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
 CC dihydrofolate reductase comprising CHO0H His6tag. The unnatural amino
 CC acid is provided exogenously. The translation system is a cell and the
 CC unnatural amino acid is biosynthesised by the cell. The present sequence
 CC represents a plasmid containing the individual genes papABC that encode
 CC the enzymes used to carry out the conversion of chorismate to the
 CC unnatural amino acid p-aminophenylalanine (pAF)
 CC
 SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;
 Query Match 95.0%; Score 296.4; DB 8; Length 3305;
 Best Local Similarity 98.4%; Pred. No. 6.6e-44;
 Matches 311; Conservative 0; Mismatches 1; Indels 4; Gaps 1;
 QY 1 ATGACCGAGCAAGACGCTGCA---GCGGCTGCGCGGAGCTCGACCGCTCGACGG 56
 DB 2990 ATGACCGAGCAAGACGCTGCAAGCTGCGGCTGCGCGGAGCTCGACCGCTCGACGG 3049
 QY 57 GAGGCTCTGGAACACGGTGGGCGCGCCGATCGAAGCTGCTCGGATCGCGCGGTACAA 116
 DB 3050 GAGGCTCTGGAACACGGTGGGCGCGCCGATCGAAGCTGCTCGGATCGCGCGGTACAA 3109
 QY 117 GTCCCGGACAGGCTCCCGATGATGACAGCCGCGCGGATCGCTGATCAAGACAGGGC 176
 DB 3110 GTCCCGGACAGGCTCCCGATGATGACAGCCGCGCGGATCGCTGATCAAGACAGGGC 3169
 QY 177 CGCCCGCTACGCGCGCGACCAAGCTGCAAGATGTTCTGTGTAACCTTACAGAGT 236
 DB 3170 CGCCCGCTACGCGCGCGACCAAGCTGCAAGATGTTCTGTGTAACCTTACAGAGT 3229
 QY 237 GATCATACAGGAGATGTCGCGCGTGAAGATCTGTGATGACCGGAGAGGCTGACAGGC 296
 DB 3230 GATCATACAGGAGATGTCGCGCGTGAAGATCTGTGATGACCGGAGAGGCTGACAGGC 3289
 QY 297 CGAGAGCCGCGCGTGA 312
 DB 3290 CGAGAGCCGCGCGTGA 3305

RESULT 4
 AB269798
 ID AB269798 standard; DNA; 12391 BP.

XX AB269798;
 AC 08-APR-2003 (first entry)
 XX
 DT Plasmid plasc-papabc.
 XX
 DE DHFR; translation: orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
 KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
 KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
 KW biosynthesis; p-aminophenylalanine; pAF.
 XX Synthetic.
 OS
 XX
 XX MO200285923-A2.
 PN 31-OCT-2002.
 PD 19-APR-2002; 2002MO-US012465.
 XX 19-APR-2001; 2001US-0285030P.
 PR 06-FEB-2002; 2002US-0355514P.
 XX (SCRI) SCRIPPS RES INST.
 PA Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
 PI Meggers EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;
 XX WPI; 2003-120430/11.
 DR
 XX Composition useful for producing protein comprising unnatural amino acid,
 PT has translation system comprising orthogonal tRNA and orthogonal
 PT aminoacyl tRNA synthetase.
 XX Example 4; Page 124-127; 188pp; English.
 PS
 XX The invention relates to a novel composition comprising a translation
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
 CC -tRNA with at least one unnatural amino acid in the translation system
 CC and the O-tRNA recognises at least one selector codon. A composition of
 CC the invention is useful for producing at least one protein comprising at
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TAG mutant of mouse
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
 CC dihydrofolate reductase comprising CHO0H His6tag. The unnatural amino
 CC acid is provided exogenously. The translation system is a cell and the
 CC unnatural amino acid is biosynthesised by the cell. The present sequence
 CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine
 CC (pAF) in vivo
 CC
 SQ Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;
 Query Match 84.9%; Score 264.8; DB 8; Length 12391;
 Best Local Similarity 97.9%; Pred. No. 2.3e-38;
 Matches 280; Conservative 0; Mismatches 2; Indels 4; Gaps 1;
 QY 1 ATGACCGAGCAAGACGCTGCA---GCGGCTGCGCGGAGCTCGACCGCTCGACGG 56
 DB 2055 ATGACCGAGCAAGACGCTGCAAGCTGCGGCTGCGCGGAGCTCGACCGCTCGACGG 2114
 QY 57 GAGGCTCTGGAACACGGTGGGCGCGCCGATCGAAGCTGCTCGGATCGCGCGGTACAA 116
 DB 2115 GAGGCTCTGGAACACGGTGGGCGCGCCGATCGAAGCTGCTCGGATCGCGCGGTACAA 2174
 QY 117 GTCCCGGACAGGCTCCCGATGATGACAGCCGCGCGGATCGCTGATCAAGACAGGGC 176
 DB 2175 GTCCCGGACAGGCTCCCGATGATGACAGCCGCGCGGATCGCTGATCAAGACAGGGC 2234
 QY 177 CGCCCGCTACGCGCGCGACCAAGCTGCAAGATGTTCTGTGTAACCTTACAGAGT 236
 DB 2235 CGCCCGCTACGCGCGCGACCAAGCTGCAAGATGTTCTGTGTAACCTTACAGAGT 2294

QY 237 GATCATCAGCGAGATGTCCCGCTCGAGACCTGTGTGATGACCCG 282
 |||||
 DB 2295 GATCATCAGCGAGATGTCCCGCTCGAGACCTGTGTGATGACCCG 2340
 |||||
 RESULT 5
 AAT58552 standard; cDNA, 387 BP.
 ID AAT58552
 AC AAT58552;
 XX
 DT 02-APR-1997 (first entry)
 XX
 DE Streptomyces pristinaespiralis papB gene.
 XX
 KM Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
 KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papB; papM; papB;
 KM papC; isomerisation; aromatisation; N-methyltransferase; ds.
 XX
 OS Streptomyces pristinaespiralis.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..387
 FT /*tag= a
 FT /product= "PapB"
 XX
 PN MO9601901-A1.
 XX
 PD 25-JAN-1996.
 XX
 PF 04-JUL-1995; 95WO-FR000889.
 XX
 PR 08-JUL-1994; 94FR-00008478.
 XX
 PA (RHON) RHONE POULENC RORER SA.
 PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
 PI Barriere J, Debussche L, Famechon A, Dutruc-Rosset G;
 XX
 DR WPI; 1996-097631/10.
 DR P-PSDB; AAW11583.
 XX
 PT New streptogramin B derivs. useful as antibiotics - produced by new
 PT mutants of Streptomyces having altered genes for streptogramin B
 PT biosynthesis.
 XX
 PS Claim 18; Page 106-107; 146pp; French.
 CC The papA and papM genes of S.pristinaespiralis are involved in the
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
 CC pristinamycin 1A. The region between these two genes was sequenced and
 CC two open reading frames were identified. The first (papC) was on the
 CC complementary strand and decodes to an amino acid sequence with homology
 CC to the region of E.coli TyrA which has been implicated in aromatisation
 CC reactions. The papC gene product is likely to be involved in a similar
 CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
 CC phenylpyruvate during DMPAPA synthesis. The second open reading frame
 CC (papB) could be decoded to give a product with homology to the region of
 CC TyrA which has chorismate mutase activity. The papB gene product is
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
 CC deoxy 4-amino prephenate acid. Disruption of the papB and papC genes can
 CC be used to produce strains of S.pristinaespiralis which are unable to
 CC produce the antibiotic pristinamycin I but which may be able to produce
 CC new, modified forms of it
 XX
 SQ Sequence 387 BP; 50 A; 187 C; 110 G; 40 T; 0 U; 0 Other;
 Query Match 43.5%; Score 135.6; DB 2; Length 387;
 Best Local Similarity 69.8%; Pred. No. 2.3e-15;
 Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
 QY 14 ACAGCTGCAGCGCGCTGCGCGGAGCTCGAGCCCTCGACGGAAGCTCTTGACACG 73
 |||||

DB 59 ACCCCCTCGACGCGCTGCGCGCCCGCTTGACGCGCGGACGCCCGCTGTGAGACCGG 118
 QY 74 TCGGAGCGCCGATCGACCTCGGTGTCCGATCGCGGATACAAAGTCCCGGACGCGGTCC 133
 |||||
 DB 119 TCCGACACGCTCGACATCTGCTCCGATCGGCGAAGTACAAAGCCCTTCACACAGTGC 178
 QY 134 CGATGATGACGCGCGCGGGGTGACCTGTGTAAGACAGGGCGCCGCTACGCGCGG 193
 |||||
 DB 179 CGATGATGACGCGCGCCCGGATGCGCCAGGTCCACGCAAGCCCGCTACGCGCGG 238
 QY 194 ACCAGCGCTCGACGAATGCTTCGTGTAACCTGACGCGGATCATCGGAGATGT 253
 |||||
 DB 239 ACCAGCGCATCGACCCCGCTTCTGCGCACCTGTACGACAGATCATCGGAGACCT 298
 QY 254 GCGCGCTCGAGACCTGTGAT 275
 |||||
 DB 299 GCCGCTCGAGACGAGTGGAT 320
 |||||
 RESULT 6
 AAT59268
 ID AAT59268 standard; cDNA; 2888 BP.
 XX
 AC AAT59268;
 XX
 DT 02-APR-1997 (first entry)
 XX
 DE Streptomyces pristinaespiralis papA and papM intergenic region.
 XX
 KM Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
 KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
 KM papC; isomerisation; aromatisation; N-methyltransferase; ds.
 XX
 OS Streptomyces pristinaespiralis.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT /*tag= a
 FT /product= "PapA"
 FT /note= "C-terminal coding region only, i.e. a partial
 FT open reading frame"
 FT complement(949..1836)
 FT /*tag= b
 FT /product= "PapC"
 FT 1873..2262
 FT /*tag= c
 FT /product= "PapB"
 FT 2259..2888
 FT /*tag= d
 FT /product= "PapM"
 FT /note= "N-terminal coding region only, i.e. a partial
 FT open reading frame"
 XX
 PN WO9601901-A1.
 XX
 PD 25-JAN-1996.
 XX
 PF 04-JUL-1995; 95WO-FR000889.
 XX
 PR 08-JUL-1994; 94FR-00008478.
 XX
 PA (RHON) RHONE POULENC RORER SA.
 PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
 PI Barriere J, Debussche L, Famechon A, Dutruc-Rosset G;
 XX
 DR WPI; 1996-097631/10.
 DR P-PSDB; AAW11582, AAW11583.
 XX
 PT New streptogramin B derivs. useful as antibiotics - produced by new
 PT mutants of Streptomyces having altered genes for streptogramin B
 PT biosynthesis.
 XX

PS Example 1; Page 102-104; 146pp; French.
 XX The papA and papM genes of *S. pristinaespiralis* are involved in the
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPPA), a precursor for
 CC pristinamycin 1A. The region between these two genes was sequenced and
 CC two open reading frames were identified. The first (papC) was on the
 CC complementary strand and decodes to an amino acid sequence with homology
 CC to the region of *E. coli* TyrA which has been implicated in aromatization
 CC reactions. The papC gene product is likely to be involved in a similar
 CC aromatization of 4-deoxy 4-amino prephenate to give 4-amino
 CC phenylpyruvate during DMPPA synthesis. The second open reading frame
 CC (papB) could be decoded to a product with homology to the region of
 CC TyrA which has chorismate mutase activity. The papB gene product is
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
 CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
 CC used to produce strains of *S. pristinaespiralis* which are unable to
 CC produce the antibiotic pristinamycin I but which may be able to produce
 CC new, modified forms of it
 XX
 SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;
 Query Match 43.5%; Score 135.6; DB 2; Length 2888;
 Best Local Similarity 69.8%; Pred. No. 1.9e-15;
 Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
 QY 14 ACCGAGCTGCAGCGCGCTGCGCGGAGCTCGAGCGCCCTCGACGGGACGCTCTCTGACACGG 73
 DB 1931 ACCCCCTCGACGCGCTGCGCGCGCGCTCGAGCGCGGAGCGCCCTCTGTGAGCGCG 1990
 QY 74 TGGCGCGCGCGCATCGACTCGGTGTCCGATCGCGCGGTACAAGTCCCGGACGCGCTCC 133
 DB 1991 TCCGACACGCGCTCGACATCTGCTGGCATCGCGCGATACAGCGCTCTCACAGGTCC 2050
 QY 134 CGATGATGCAGCGCGCGCGGTCACTGTGTCAAGACAGGCGCGCGCTACGCGCGCG 193
 DB 2051 CGATGATGCAGCGCGCGCGGTCACTGTGTCAAGACAGGCGCGCGCTACGCGCGCG 2110
 QY 194 ACCAGCGCGCTCGAGCATCTCTCTGTGGAACCTCTTCAGACGTATATATACAGAGATGT 253
 DB 2111 ACCAGCGCGCATCGACCGCGCTCTCTGTGCGACCTGTATGACATATATACAGAGACT 2170
 QY 254 GCGCGCTCGAGGACCTGTGTAT 275
 DB 2171 GCGCGCTCGAGGACGAGTGTAT 2192

RESULT 7
 ABZ66707 standard; DNA: 1458 BP.
 XX
 AC ABZ66707;
 XX
 DT 21-MAR-2003 (first entry)
 XX
 DE Orthosomycin biosynthetic polynucleotide SEQ ID NO 76.
 XX
 KW Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.
 XX
 OS Streptomyces mobaraensis.
 XX
 PN WO200279505-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-CA000432.
 XX
 PR 28-MAR-2001; 2001US-0279095P.
 XX
 PR 30-MAR-2001; 2001US-0279709P.
 XX
 PR 20-APR-2001; 2001US-0285214P.
 XX
 PA (ECOP-) ECOPIA BIOSCIENCES INC.
 XX
 PI Farnet CM, Zazopoulos E, Staffa A;

XX
 DR WPI: 2003-058435/05.
 DR P-PSDB; ABP99244.
 XX
 PT Identifying orthosomycin biosynthetic gene, gene fragment or gene
 PT cluster, by detecting presence of nucleic acid sequence corresponding to
 PT 17 of flambamycins protein families.
 XX
 PS Example 2; Page 231-232; 51pp; English.
 XX
 CC The invention relates to identifying orthosomycin biosynthetic genes and
 CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
 CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
 CC ABP99362). The method is useful for identifying an orthosomycin
 CC biosynthetic gene, gene fragment or gene cluster, especially an
 CC evernimycin-type or avilamycin-type orthosomycin biosynthetic gene,
 CC gene fragment or gene cluster. The method is useful for detecting the
 CC presence of any organism that contains DNA for the production of
 CC orthosomycins (both evernimycin-type orthosomycins and avilamycin-type
 CC orthosomycins) regardless of the level at which genes for orthosomycin
 CC production are expressed by the organism or the amount of orthosomycin
 CC produced by the organism. This allows for the detection of new
 CC orthosomycin natural products, not produced by the organism
 XX
 SQ Sequence 1458 BP; 229 A; 530 C; 491 G; 208 T; 0 U; 0 Other;
 Query Match 17.6%; Score 54.8; DB 10; Length 1458;
 Best Local Similarity 49.0%; Pred. No. 0.41;
 Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
 QY 6 CGAGCAGAGACGCTGAGCGGCTGCGCGCGCTCGAGCGCCCTCGACGGGACGCTCTCT 65
 DB 684 CACACAGAGACGCTCTCTGAGCGCGCAGAGAGACGAGACTGAGCGAGTCTGCG 743
 QY 66 GGAACAGGTGCGCGCGCGCATCGACTCGGTGTCCGATCGCGCGGTACAAGTCCCGGCA 125
 DB 744 CTGCGTGTCTCAAGAGAACCCGACCGTCCGCGACCTCATGGCCGAGCGGCGAGTACC 803
 QY 126 CGCGGTCCGATGATGAGCGCGCGCGGTGACCTGTGTCAAGACAGGCGCGCGCTA 185
 DB 804 GCGCGTCCGCGGTGAGACGAGACTTCTCTGCTGCGCGACGACTTCCACGCGCGCGCTA 863
 QY 186 CGCGCGCGACGACGCGCTCGACGATCTCTCTGTGTGAACTCTACGACCTGATCATCAG 245
 DB 864 CTACTTGGCGCGGAGCGCGCTCTCTCTGCTCGACCGCGTGTCTCAACCGCGCGCTCA 923
 QY 246 GGAATGTGCGCGGTGAGGACCTGTGTATGAGCGCGGAGAGGCTGACGCGCGAGAC 303
 DB 924 GCGCGTCTACAGCGGAGCTGTGCGCGCGAGCTCGTGTGCTGCGCACGAGTGAACAGAGAC 981

RESULT 8
 ABZ66808/c
 ID ABZ66808 standard; DNA: 45055 BP.
 XX
 AC ABZ66808;
 XX
 DT 21-MAR-2003 (first entry)
 XX
 DE Orthosomycin biosynthetic gene cluster SEQ ID NO 277.
 XX
 KW Orthosomycin; biosynthesis; evernimycin; avilamycin; gene; ds.
 XX
 OS Streptomyces mobaraensis.
 XX
 PN WO200279505-A2.
 XX
 PD 10-OCT-2002.
 XX
 PF 28-MAR-2002; 2002WO-CA000432.
 XX
 PR 28-MAR-2001; 2001US-0279095P.
 XX
 PR 30-MAR-2001; 2001US-0279709P.

```
PR 20-APR-2001; 2001US-0285214P.
XX
XX (ECOP-) ECOPIA BIOSCIENCES INC.
XX
XX Farnet CM, Zazopoulos E, Staffa A;
XX WPI; 2003-058435/05.
XX
XX Identifying orthosomycin biosynthetic gene, gene fragment or gene
XX cluster, by detecting presence of nucleic acid sequence corresponding to
XX 17 of flamamycins protein families.
XX
PS Example 2; Page 425-450; 511pp; English.
XX
XX The invention relates to identifying orthosomycin biosynthetic genes and
XX its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the
XX presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
XX ABP99362). The method is useful for identifying an orthosomycin
XX biosynthetic gene, gene fragment or gene cluster, especially an
XX evernimicin-type or avilamycin-type orthosomycin biosynthetic gene,
XX gene fragment or gene cluster. The method is useful for detecting the
XX presence of any organism that contains DNA for the production of
XX orthosomycins (both evernimicin-type orthosomycins and avilamycin-type
XX orthosomycins) regardless of the level at which genes for orthosomycin
XX production are expressed by the organism or the amount of orthosomycin
XX produced by the organism. This allows for the detection of new
XX orthosomycin natural products, not produced by the organism
XX
SQ Sequence 45055 BP; 6228 A; 16576 C; 16043 G; 6208 T; 0 U; 0 Other;
XX
XX Query Match 17.6%; Score 54.8; DB 10; Length 45055;
XX Best Local Similarity 49.0%; Pred. No. 0.3;
XX Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
XX
QY 6 CGAGCAGACGACCTGACGCGCTGCGCGGAGCTGACGCGCTGACGCGGACGCTCCT 65
DB 37629 CACCCAGAACCGCTTCTGAGCGCCGCAAGACGACGACTGAGCGCATGCTCGC 37570
QY 66 GGACACGATGCGCGCGCGCATCGACCTCGGATGTCGCGCATCGCGGTACAAAGTCCCGCA 125
DB 37569 CTGCGTGTCTCCAGAGAACCCGACGATCGCGCATCTCATGCGGAGGCGAAGTACCAAGC 37510
QY 126 CGGCGTCCGATGATGACGCGCGCGGATGACGCTGATGATGAGACAGGCGCGCGCTTA 185
DB 37509 GGGCGCTCCGAGTGAAGAGGACTTCTCTACGTCGCGGACGCTTCCACGCGCCCGGCTTA 37450
QY 186 CGCGCGGACCAAGCGGCTGACGAATCGTTCCTGATGAACTCTACAGAGTGATCATAC 245
DB 37449 CTACCTGCGGAGGAGCGCGCTGCTTCTGACCGCGCTGCTCAACCGCGCTCACT 37390
QY 246 GGAGATGTGCGCGCTGAGAGACCTGATGATGAGCCGCGAGAGCCTGACGCGCGAGGAC 303
DB 37389 GGGCGCTCTACAGCGGAGATGTGTGCGCGGAGAGTCCGATGCTGCCACGCTGAACGAGAC 37332
XX
XX RESULT 9
XX ABD12889
XX ID ABD12889 standard; DNA; 507 BP.
XX
XX ABD12889;
XX
XX 29-UTL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polynucleotide #11493.
XX
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
XX antibacterial.
XX
XX Pseudomonas aeruginosa.
XX
XX US6511795-B1.
XX
XX 22-APR-2003.
XX
XX
```

```
XX
XX 18-FEB-1999; 99US-00252991.
XX
XX PF
XX 18-FEB-1998; 98US-0074788P.
XX
XX PR 27-UTL-1998; 98US-0094190P.
XX
XX (GENO-) GENOME THERAPEUTICS CORP.
XX
XX PI Rubenfield MJ, Nolling J, DeLoughery C, Bush D;
XX WPI; 2003-615109/58.
XX
XX DR P-PSDB; AB079318.
XX
XX PT Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
XX PS Disclosure; SEQ ID NO 11493; 455pp; English.
XX
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a P. aeruginosa nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-P. aeruginosa drugs, as templates for recombinant
XX production of P. aeruginosa-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of P. aeruginosa-caused
XX infection, and in detection of P. aeruginosa sequences or other sequences
XX of Pseudomonas species using bioclip technology. Sequences ABD01397-
XX ABD17967 represent P. aeruginosa polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
SQ Sequence 507 BP; 86 A; 176 C; 163 G; 82 T; 0 U; 0 Other;
XX
XX Query Match 17.5%; Score 54.6; DB 11; Length 507;
XX Best Local Similarity 49.8%; Pred. No. 0.49;
XX Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
XX
QY 33 CGGAGCTTCGACGCGCTTGAAGGAGCGCTCTGACACGATGCGGCGCGCATGCACT 92
DB 149 CGAAGGTTCTTCTGAGACGCGCTGCGCGCGCATGATGATGATGCGGCGCGCATAT 208
QY 93 CGGTGCCCATCGCGCGCTTACAAATCCCGGACAGCGCGCTCCGATGATGACCGCGCGG 152
DB 209 CGCGTGAAGTTGCTTCATATCTTCAACGAGGCTGCGCGCGCGAGACCACTTGTCTTACCG 268
QY 153 GGTACACCTGATCAAGAGAGGCGCGCGCTTACAGCGCGCGCGACAGCGCGCTGACAGATC 212
DB 269 TCGGACCTGTTCTGCGGCGCTTGAACCGGACGCTCGCGAGCATCTCGCGGACGAAT 328
QY 213 GTTCTGTGAACCTTACAGCTGATCATCAAGAGATGTGCGCGGTGAGAGACTGT 272
DB 329 GGGCAAGAGAGGCGCTGACCTGCAATTCATCAAGAGCATGCGCGGATGCAAAACAGCG 388
QY 273 GATGAGCGCGGAGAGCTTACGCGCGGAGAGACCGCGCGG 309
DB 389 CGATGGACGCTGCGCGGAGACCTTCAAGAGAGCGCGG 425
XX
XX RESULT 10
XX ABD12819/C
XX ID ABD12819 standard; DNA; 1359 BP.
XX
XX ABD12819;
XX
XX 29-UTL-2004 (first entry)
XX
XX Pseudomonas aeruginosa polynucleotide #11423.
XX
XX
```


KM Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
KM antibacterial.
XX
OS *Pseudomonas aeruginosa*.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO79248.
XX
XX
PT Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 11423; 455pp; English.
XX
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a *P. aeruginosa* nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-*P. aeruginosa* drugs, as templates for recombinant
XX production of *P. aeruginosa*-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of *P. aeruginosa*-caused
XX infection, and in detection of *P. aeruginosa* sequences or other sequences
XX of *Pseudomonas* species using biochip technology. Sequences ABD01397-
XX ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1359 BP; 212 A; 463 C; 446 G; 238 T; 0 U; 0 Other;
Query Match 17.5%; Score 54.6; DB 11; Length 1359;
Best Local Similarity 49.8%; Pred. No. 0.45;
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 33 CGCGAGCTCGAGCGCTTCGACGGAGCGCTTCGACACGGTGGCGCGCATCGACT 92
Db 928 CGAAGCGTTCTTCTCGAGCGCTTCGCGCGCGCATCGTGTGCGCGCGCATCAT 869
QY 93 CGGTGTCGATCGCGCGGTACAGTCCCGGACGGGCTCCGATGATGACGCCGCGC 152
Db 868 CGCGGTGAGTTCCTCGATCTTCAACGGGCTTCGCGCGCGACCACTTGCTCTACCG 809
QY 153 GGTACGCTGTCGACAGGAGCGCGCGCTACGCGCGACGACGCGCTTCGACGATC 212
Db 808 TCGGACCTGTTCTTCGCGCGCTTCGACCGGAGCTTCGCGAGATCTGCGCGCGAATT 749
QY 213 GTTCTGTTGTAACCTTACGACGTGATCATCAAGAGATGTCCGCTCGAGACTGT 272
Db 748 GGGCAAGAGGCGCTGACCTGACCTGACGATTCACAGGACATCGCCGTATCGAACAAGCGC 689
QY 273 GATGAGCGGGAGAGCTGACGCGCGGACGCGCGCGC 309
Db 688 CGATGGACCTTGGCGCGACCTTCAAGAGAGCGCGC 652

RESULT 11
ABD12866

ID ABD12866 standard; DNA; 1371 BP.
XX
AC ABD12866;
XX
DT 29-JUL-2004 (first entry)
XX
DE *Pseudomonas aeruginosa* polynucleotide #11470.
XX
KM Bacterial infection; gene; ds; *Pseudomonas aeruginosa* infection;
KM antibacterial.
XX
OS *Pseudomonas aeruginosa*.
XX
PN US6551795-B1.
XX
PD 22-APR-2003.
XX
PF 18-FEB-1999; 99US-00252991.
XX
PR 18-FEB-1998; 98US-0074788P.
PR 27-JUL-1998; 98US-0094190P.
XX
PA (GENO-) GENOME THERAPEUTICS CORP.
XX
PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX
DR WPI; 2003-615309/58.
DR P-PSDB; ABO79295.
XX
XX Novel isolated nucleic acid encoding *Pseudomonas aeruginosa* polypeptide,
XX useful as molecular targets for diagnostics, prophylaxis and treatment of
XX pathological conditions resulting from bacterial infection.
XX
PS Disclosure; SEQ ID NO 11470; 455pp; English.
XX
XX The invention relates to *Pseudomonas aeruginosa* polypeptides and the
XX polynucleotides encoding them. The sequences are useful in diagnosis and
XX therapy of pathological conditions, as molecular targets for diagnostics,
XX prophylaxis and treatment of pathological conditions resulting from a
XX bacterial infection, for evaluating a compound, such as a polypeptide,
XX for the ability to bind a *P. aeruginosa* nucleic acid, as components of
XX effective antibacterial targets, as targets for antibacterial drugs,
XX including anti-*P. aeruginosa* drugs, as templates for recombinant
XX production of *P. aeruginosa*-derived peptides or polypeptides, as target
XX components for diagnosis and/or treatment of *P. aeruginosa*-caused
XX infection, and in detection of *P. aeruginosa* sequences or other sequences
XX of *Pseudomonas* species using biochip technology. Sequences ABD01397-
XX ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:
XX The sequence data for this patent did not form part of the printed
XX specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
SQ Sequence 1371 BP; 241 A; 445 C; 464 G; 421 T; 0 U; 0 Other;
Query Match 17.5%; Score 54.6; DB 11; Length 1371;
Best Local Similarity 49.8%; Pred. No. 0.45;
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 33 CGCGAGCTCGAGCGCTTCGACGGAGCGCTTCGACACGGTGGCGCGCATCGACT 92
Db 486 CGAAGCGTTCTTCTCGAGCGCTTCGCGCGCGCATCGTGTGCGCGCGCATCAT 545
QY 93 CGGTGTCGATCGCGCGGTACAGTCCCGGACGGGCTCCGATGATGACGCCGCGC 152
Db 486 CGCGGTGAGTTCCTCGATCTTCAACGGGCTTCGCGCGCGACCACTTGCTCTACCG 605
QY 153 GGTACGCTGTCGACAGGAGCGCGCGCTACGCGCGACGACGCGCTTCGACGATC 212
Db 606 TCGGACCTGTTCTTCGCGCGCTTCGACCGGAGCTTCGCGAGATCTGCGCGCGAATT 665
QY 213 GTTCTGTTGTAACCTTACGACGTGATCATCAAGAGATGTCCGCTCGAGACTGT 272
Db 666 GGGCAAGAGGCGCTGACCTGACGATTCACAGGACATCGCCGTATCGAACAAGCGC 725

OY		273 GATGAGCCGGAGAGCCTTAACGGCGCAGACGCCGC9	309
Dd		726 CGATGGCAACCCTGGCCGCCACCTCTCAAAGAAGCGCCG9	762
<hr/>			
	RESULT 12		
	AD873531	standard; cDNA; 73882 BP.	
XX	AD873531;		
AC			
XX			
DT	02-DEC-2004	(first entry)	
XX			
DE	tcp gene cluster.		
XX			
KM	ss; gene; tcp ; gene cluster ; biosynthesis ; glycopeptide ; teicoplanin;		
KW	antibiotic; bal; cep; com ; sta ; 4-hydroxy-phenylglycine ; 3:		
KM	5-dihydroxy-phenylglycine; heptapeptide skeleton ; cross-linking ;		
KM	aromatic residue; N-acyl-beta-D-glucosamine ; core structure ; export;		
XX	resistance; chlorination ; alpha-hydroxylation ; mannose attachment.		
OS	Actinoplanes teichomyceticus; ATCC31121.		
FH			
Key	Location/Qualifiers		
CDS	652..1395		
FT	/tag= a		
FT	/label= ORF1		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	/product= "MurP"		
FT	2017..3234		
CDS	/tag= b		
FT	/label= ORF2		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	3227..4264		
FT	/tag= c		
FT	/label= ORF3		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	4264..4872		
FT	/tag= d		
FT	/label= ORF4		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	complement(5634)..6173)		
CDS	/tag= e		
FT	/label= ORF5		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	complement(6286)..7377)		
CDS	/tag= f		
FT	/label= ORF6		
FT	/note= "Required for export of teicoplanin outside of the cytoplasm," and for conferring resistance to teicoplanin"		
FT	complement(7385)..8065)		
FT	/tag= g		
FT	/label= ORF7		
FT	/note= "Required for regulating expression of one or more of the genes of tcp gene cluster"		
FT	8924..10105		
CDS	/tag= h		
FT	/label= ORF8		
FT	/note= "Required for the addition and formation of N-acyl-beta-D-glucosamine residues"		
FT	10597..116827		
FT	/tag= i		
FT	/label= ORF9		
FT	/note= "Required for synthesis of the heptapeptide skeleton"		

FT	CDS	16824..20009
FT		/*tag= j
FT		/label= ORF10
FT		/note= "Required for synthesis of the heptapeptide
FT		skeleton"
FT	CDS	20053..32256
FT		/*tag= k
FT		/label= ORF11
FT		/note= "Required for synthesis of the heptapeptide
FT		skeleton"
FT	CDS	32276..37873
FT		/*tag= l
FT		/label= ORF12
FT		/note= "Required for synthesis of the heptapeptide
FT		skeleton"
FT	CDS	37886..38095
FT		/*tag= m
FT		/label= ORF13
FT	CDS	38401..39222
FT		/*tag= n
FT		/label= ORF14
FT	CDS	39268..41049
FT		/*tag= o
FT		/label= ORF15
FT		/note= "Required for attachment of mannosyl residue"
FT	CDS	41343..43289
FT		/*tag= p
FT		/label= ORF16
FT		/note= "Required for export of telcoplanin outside of the
FT		cytoplasm, and for conferring resistance to telcoplanin"
FT	CDS	43373..43585
FT		/*tag= q
FT		/label= ORF17
FT	CDS	43696..44871
FT		/*tag= r
FT		/label= ORF18
FT		/note= "Required for cross-linking of aromatic residues
FT		at position 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT		telcoplanin"
FT	CDS	44894..46048
FT		/*tag= s
FT		/label= ORF19
FT		/note= "Required for cross-linking of aromatic residues
FT		at position 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT		telcoplanin"
FT	CDS	46038..47234
FT		/*tag= t
FT		/label= ORF20
FT		/note= "Required for cross-linking of aromatic residues
FT		at position 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT		telcoplanin"
FT	CDS	47484..49004
FT		/*tag= u
FT		/label= ORF21
FT		/note= "Required for the chlorination of the aromatic
FT		residues"
FT	CDS	49166..50344
FT		/*tag= v
FT		/label= ORF22
FT		/note= "Required for cross-linking of aromatic residues
FT		at position 2 and 4, 4 and 6, 1 and 3, and 5 and 7 of
FT		telcoplanin"
FT	CDS	50461..51687
FT		/*tag= w
FT		/label= ORF23
FT		/note= "Required for the addition and formation of N-acyl
FT		-beta-glucosamine residues"
FT	CDS	51833..52804
FT		/*tag= x
FT		/label= ORF24
FT		/note= "Required for the addition and formation of N-acyl
FT		-beta-glucosamine residues"
FT	CDS	52948..54540

```

FT      /*cag= y
FT      /label= ORF25
FT      /note= "Required for the beta-hydroxylation of the Tyr
FT      residues"
FT      54757. .56553
FT      /*cag= z
FT      /label= ORF26
FT      /note= "Required for the addition and formation of N-acyl
FT      -beta-glucosamine residues"
FT      56924. .57979
FT      CDS
FT      /*cag= aa
FT      /label= ORF27
FT      58746. .59735
FT      CDS
FT      /*cag= ab
FT      /label= ORF28
FT      /note= "Required for regulating expression of one or more
FT      of the genes of tcp gene cluster"
FT      60477. .62855
FT      CDS
FT      /*cag= ac
FT      /label= ORF29
FT      /note= "Required for regulating expression of one or more
FT      of the genes of tcp gene cluster"
FT      63036. .64157
FT      CDS
FT      /*cag= ad
FT      /label= ORF30
FT      /note= "Required for synthesis of the 3,5-dihydroxy-
FT      phenylglycine residues"
FT      64154. .64819
FT      CDS
FT      /*cag= ae
FT      /label= ORF31
FT      /note= "Required for synthesis of the 3,5-dihydroxy-
FT      phenylglycine residues"
FT      64816. .66117
FT      CDS
FT      /*cag= af
FT      /label= ORF32
FT      /note= "Required for synthesis of the 3,5-dihydroxy-
FT      phenylglycine residues"
FT      66146. .66952
FT      CDS
FT      /*cag= ag
FT      /label= ORF33
FT      /note= "Required for synthesis of the 3,5-dihydroxy-
FT      phenylglycine residues"
FT      66991. .68352
FT      CDS
FT      /*cag= ah
FT      /label= ORF34
FT      /note= "Required for export of teicoplanin outside of the
FT      cytoplasm, and for conferring resistance to teicoplanin"
FT      68349. .68945
FT      CDS
FT      /*cag= ai
FT      /label= ORF35
FT      69083. .70180
FT      CDS
FT      /*cag= aj
FT      /label= ORF36
FT      /note= "Required for synthesis of the 4-hydroxy-
FT      phenylglycine residues, and required for synthesis of the
FT      3,5-dihydroxy-phenylglycine residues"
FT      70266. .71321
FT      CDS
FT      /*cag= ak
FT      /label= ORF37
FT      /note= "Required for synthesis of the 4-hydroxy-
FT      phenylglycine residues"
FT      71318. .72412
FT      CDS

Query Match      17.5%; Score 54.6; DB 13; Length 73882;
Best Local Similarity 48.5%; Pred. No. 0.31;
Matches 150; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

OY      1 ATGACCGAGCAAGCAAGCTGCGCGGCTGCGCGGAGCTTCGAGCGCTTCGAGCGGAGC 60
DB      29713 ATGACCGCGCGCGATCGCGCGCGGCGGCTTCGCGGCTCGACGAGCTGAGCGCGG 29772
OY      61 CTCCTGAGACAGGTGCGCGCGCGCATCGACTCGGTGCTCCGCAATCGCGCGGTACAAGTCC 120

```

```

DB      29773 CGCCTGTGGCTTACCCGCCGCGACGCGGTCAAGATGAACCCCTTCAGAGGATC 29832
OY      121 CGGACGCGCGTCCCGATGATGACGCCGCGCGGTGACCTGTTCAGAGACAGGCGGCC 180
DB      29833 TGGTCGTGTGCTGTGGCCCGCGCTGACGCGCGGAGCAGCATGTGCTTCGCTCCGTG 29892
OY      181 CGCTACGCGCGCGACCGACCGGCTCGACGATCGTTCGTGTAACCTCTGACAGCTGATC 240
DB      29893 GCCGCGCGCGCCCGCGAGATCCCGGCTGGGTCCGTGGTCGGGCTGTTCATGACACG 29952
OY      241 ATCAGGAGATGTGCGCGCTCGAGAGACCTGTGATGAGCCGCGAGACCTGACGCGCGAG 300
DB      29953 CTGCGCGGTCCGGGTGGCGCTCGACGCGCGCGAACCCTTCGTGACATGCTGACGACCTG 30012
OY      301 GACCGCGCG 309
DB      30013 CAGCGCGCG 30021

RESULT 13
AA064204
ID      AA064204 standard; cDNA; 1208 BP.
XX
AC      AA064204;
XX
DT      25-MAR-2003 (revised)
DT      18-NOV-1994 (first entry)
XX
DE      snac gene encoding enzyme in streptogramin biosynthetic pathway.
XX
KW      Antibiotic; streptogramin; snac; snab; snac; biosynthesis; enzyme;
KW      biosynthetic pathway; Streptomyces pristinaespiralis; ds.
XX
OS      Streptomyces pristinaespiralis.
XX
FH      Key Location/Qualifiers
FT      CDS 1..1209
FT      /*cag= a
XX
PN      FR2696189-A1.
XX
PD      01-APR-1994.
XX
PF      25-SEP-1992; 92FR-00011441.
XX
PR      25-SEP-1992; 92FR-00011441.
XX
PA      (RHON ) RHONE POULENC RORER SA.
XX
PI      Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P, Thibaut D,
PI      Zagorec M;
XX
DR      WP1; 1994-128286/16.
DR      P-PSDB; AARS4204.
XX
PT      DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bio
PT      -conversion of streptogramin(s) or prodn. of streptogramin intermediates,
PT      derivs. or hybrid antibiotics.
XX
PS      Claim 2; Page 54-55; 83pp; French.
XX
CC      The snac gene product is involved in the biosynthesis of streptogramins,
CC      antibiotics active against Gram-positive bacteria. The identification of
CC      the sequences encoding the enzymes involved in the biosynthetic pathway
CC      means that they can be isolated and manipulated. Mutant microorganisms in
CC      which a step in the streptogramin biosynthetic pathway is blocked can be
CC      cultured to produce streptogramin intermediates, which may later be
CC      converted to streptogramin derivatives. Recombinant cells may also be
CC      used for the bioconversion of streptogramins from one form to another or
CC      for the production of hybrid antibiotics. (Updated on 25-MAR-2003 to
CC      correct PN field.)
XX
SQ      Sequence 1208 BP; 190 A; 500 C; 360 G; 158 T; 0 U; 0 Other;

```

Query Match 17.3%; Score 54; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.58;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCAGACCCCTTCGACGGGACGCTCTTGACACAGGTGCGCGCCGATGACCTGGTG 97
DB 860 AGGTTCAGCCCTTCGCGCGCGCTGACGATGCGTGGTGGCCAGAAAGTGTGCGCGCG 919
QY 98 TCCGATCGCGCGGTACAATCCCGGACGCGCTCCGATGATGACGCCGCGCGGTCA 157
DB 920 GCTTCGCTTCCTCCGCTGCGAGGTCAAGTCCCTTACGCTACGCGCAAGCGCGGTG 979
QY 158 GCGTGTCAAGACAGGGCGCGCGCTACGCGCGCGACGCGCTTCGACGAATGCTTCC 217
DB 980 GCTGTTCGTGAGAGCTTGGGACCGCGCACCGTCCGCCAGAGCGCATGAGAAAGGCCA 1039
QY 218 TGGTGAACCTTCACGAGTATCATCAGGAGATGTGCCGCTGACGACCTGTGTATGA 277
DB 1040 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCATTCATCCGCGACCTGTGTGCGGC 1099
QY 278 GCGGGAGAGCTGACGCGCGCGAGACCGGC 307
DB 1100 CCATCTACGCGCGCGCGCGCTACGCGC 1129

RESULT 14

AA064201
ID AA064201 standard; cDNA; 5392 BP.

AC AA064201;
XX
DT 25-MAR-2003 (revised)
DT 18-NOV-1994 (first entry)

DE Sequence comprising the *snaA*, *snaB* and *snaC* gene cluster.

XX Antibiotic; streptogramin; *snaA*; *snaB*; *snaC*; biosynthesis; enzyme;
KM biosynthetic pathway; Streptomyces pristinaespiralis; ds.

XX Streptomyces pristinaespiralis.

XX FR2696189-A1.

XX 01-APR-1994.

XX 25-SEP-1992; 92FR-00011441.

XX 25-SEP-1992; 92FR-00011441.

XX (RHON) RHONE-POULENC RORER SA.

XX Blanc V, Blanche F, Crouzet J, Jacques N, Lacroix P, Thibaut D;
PI Zagorec M;

XX WPI; 1994-128286/16.

XX DNA involved in streptogramin antibiotic biosynthesis - for prodn. or bio
PT -conversion of streptogramin(s) or prodn. of streptogramin intermediates,
PT derivs. or hybrid antibiotics.

XX Disclosure; Page 44-47; 83pp; French.

XX This sequence comprises the *snaA*, *snaB* and *snaC* genes which are involved
CC in the biosynthesis of streptogramins, antibiotics active against Gram-
CC positive bacteria. The identification of the sequences encoding the
CC enzymes involved in the biosynthetic pathway means that they can be
CC isolated and manipulated. Mutant microorganisms in which a step in the
CC streptogramin biosynthetic pathway is blocked can be cultured to produce
CC streptogramin intermediates, which may later be converted to
CC streptogramin derivatives. Recombinant cells may also be used for the
CC bioconversion of streptogramins from one form to another or for the
CC production of hybrid antibiotics. (Updated on 25-MAR-2003 to correct PN

CC field.)
XX Sequence 5392 BP; 811 A; 2161 C; 1671 G; 749 T; 0 U; 0 Other;
SQ

Query Match 17.3%; Score 54; DB 2; Length 5392;
Best Local Similarity 50.0%; Pred. No. 0.51;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCAGACCCCTTCGACGGGACGCTCTTGACACAGGTGCGCGCGCGCATGACCTGGTG 97
DB 4417 AGGTTCAGCCCTTCGCGCGCGGTACGCGATGCGTGGTCCGCAAGAAAGTGTGCGCGCG 4476
QY 98 TCCGATCGCGCGGTACAATCCCGGACGCGCTCCGATGATGACGCCGCGCGGTCA 157
DB 4477 GCTTCGCTTCCTCCGCTGCGAGGTCAAGTCCCTTACGCCCATTCGCAAGGCGCGGTG 4536
QY 158 GCGTGTCAAGACAGGGCGCGCGCTACGCGCGCGACGCGCTTCGACGAATGCTTCC 217
DB 4537 GCTGTTCGTGAGAGGTTCGGACCGGACCGTCCGCGAGAGCGCATGAGAAAGGCCA 4596
QY 218 TGGTGAACCTTCACGAGTATCATCAGGAGATGTGCCGCTGACGACCTGTGTATGA 277
DB 4597 TCACCGAGGTCTTCGACCTGCGCGCGCGCGCGCATTCGCGACCTGTGTGCGGC 4656
QY 278 GCGGGAGAGCTGACGCGCGCGAGACCGGC 307
DB 4657 CCATCTACGCGCGCGCGCGCTACGCGC 4686

RESULT 15

AB266725
ID AB266725 standard; DNA; 837 BP.

XX AB266725;

XX 21-MAR-2003 (first entry)

XX Orthosomycin biosynthetic polynucleotide SEQ ID NO 112.

XX Orthosomycin; biosynthesis; evernimomycin; avilamycin; gene; ds.

XX Streptomyces mobaraensis.

XX WO200279505-A2.

XX 10-OCT-2002.

XX 28-MAR-2002; 2002WO-CA000432.

XX 28-MAR-2001; 2001US-0279095P.

XX 30-MAR-2001; 2001US-027909P.

XX 20-APR-2001; 2001US-0285214P.

XX (ECOP) ECOPIA BIOSCIENCES INC.

XX Farnet CM, Zazopoulos E, Staffa A;

XX WPI; 2003-058435/05.

XX P-PDB; ABP99262.

XX Identifying orthosomycin biosynthetic gene, gene fragment or gene
PT cluster, by detecting presence of nucleic acid sequence corresponding to
PT 17 of flambamycins protein families.

XX Claim 12; Page 266; 51pp; English.

XX The invention relates to identifying orthosomycin biosynthetic genes and
CC its fragment/gene cluster (AB266670-AB2666813), comprising detecting the
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-
CC ABP99362). The method is useful for identifying an orthosomycin
CC biosynthetic gene, gene fragment or gene cluster, especially an
CC evernimomycin-type or avilamycin-type orthosomycin biosynthetic gene,
CC gene fragment or gene cluster. The method is useful for detecting the

CC presence of any organism that contains DNA for the production of
 CC orthosomycins (both everimycin-type orthosomycins and avilamycin-type
 CC orthosomycins) regardless of the level at which genes for orthosomycin
 CC production are expressed by the organism or the amount of orthosomycin
 CC produced by the organism. This allows for the detection of new
 CC orthosomycin natural products, not produced by the organism
 XX

Sequence 837 BP; 105 A; 329 G; 308 G; 95 T; 0 U; 0 Other;

Query Match 17.2%; Score 53.6; DB 10; Length 837;

Best Local Similarity 52.1%; Pred. No. 0.71;

Matches 147; Conservative 0; Mismatches 129; Indels 6; Gaps 1;

Qy	28	CTGGGCGGAGAGCTCGAGCGCCCTCGAGCGGACGCTCTGACACAGGTGCGGCGCGCATC	87
Db	151	CTGCTGACCTCCGCGACGCGCTGGCCGAGGAGCTCCGCGGCGCTCCGCGCGCTGCTC	210
Qy	88	GACCTCGGTGTCGCGATCGCGGCTACAGTCCCGGACGCGCGCTCCGATGATGACGCC	147
Db	211	GACTTGGCGCGCGGACGTGCGCTTACCGCGCGCTGCTCCGCGCGCGCGAGCTGAGACC	270
Qy	148	GCGCGGCTCAGCTGTGCAAGACAGGCGCGCGCTACCGCGCGGACCAAGCGGCTGAC	207
Db	271	GCGGAGATGCGCGCGCGGCGGAGACTGACCGCGGACGAGCTGAGACGCGGACGCGCTG	330
Qy	208	GAATGCTTCTGTGGAACCTCTACGAGCTGATCATACGAGATGTGCGGCTGAGAC	267
Db	331	AGGCGCTGCGCGAGCGGCTGCTGACGCGATCTCTC-----CACCAGGTCCTGAG	384
Qy	268	CTGGTATGAGCGCGGAGAGCTTGACCGCGGAGACCGGCGG	309
Db	385	CACGTGACGAGCCGAGACCCCACTTGGGAGGCGCTCCGG	426

Search completed: October 4, 2005, 21:38:06
 Job time : 191.488 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 21:04:23 ; Search time 57.8815 Seconds
(without alignments)
8820.059 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312
Sequence: 1 atcgccgagagagacgact.....cgcgccgagagcgcgcgtga 312

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued Patents NA: *
1: /cgn2_6/pdata/1/ina/5A COMB.seq: *
2: /cgn2_6/pdata/1/ina/5B COMB.seq: *
3: /cgn2_6/pdata/1/ina/6A COMB.seq: *
4: /cgn2_6/pdata/1/ina/6B COMB.seq: *
5: /cgn2_6/pdata/1/ina/PCUS COMB.seq: *
6: /cgn2_6/pdata/1/ina/backfile1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	135.6	43.5	387	3 US-08-765-907A-4	Sequence 4, Appli
2	135.6	43.5	387	4 US-09-987-614A-4	Sequence 4, Appli
3	135.6	43.5	2888	3 US-08-765-907A-1	Sequence 1, Appli
4	135.6	43.5	2888	4 US-09-987-614A-1	Sequence 1, Appli
5	54.6	17.5	507	4 US-09-252-991A-11493	Sequence 11493, A
6	54.6	17.5	1359	4 US-09-252-991A-11423	Sequence 11423, A
7	54.6	17.5	1371	4 US-09-252-991A-11470	Sequence 11470, A
8	54.2	17.4	1963	4 US-09-902-540-2740	Sequence 2740, Ap
9	54.2	17.4	15689	4 US-09-902-540-1129	Sequence 1129, Ap
10	54	17.3	1208	2 US-08-403-852D-4	Sequence 4, Appli
11	54	17.3	1208	3 US-08-510-646B-4	Sequence 4, Appli
12	54	17.3	1208	3 US-09-231-818-4	Sequence 4, Appli
13	54	17.3	1208	3 US-09-635-359B-4	Sequence 4, Appli
14	54	17.3	5392	2 US-08-403-852D-1	Sequence 1, Appli
15	54	17.3	5392	3 US-08-510-646B-1	Sequence 1, Appli
16	54	17.3	5392	3 US-09-231-818-1	Sequence 1, Appli
17	54	17.3	5392	4 US-09-635-359B-1	Sequence 1, Appli
18	52.8	16.9	510	4 US-09-252-991A-8175	Sequence 8175, Ap
19	52.8	16.9	1116	4 US-09-252-991A-8278	Sequence 8278, Ap
20	52.8	16.9	1125	4 US-09-252-991A-8074	Sequence 8074, Ap
21	51.2	16.4	1194	4 US-09-902-540-9323	Sequence 9323, Ap
22	51.2	16.4	6816	3 US-09-404-650-1	Sequence 1, Appli
23	51.2	16.4	6816	4 US-09-935-541-1	Sequence 1, Appli
24	51.2	16.4	6855	3 US-09-404-650-3	Sequence 3, Appli
25	51.2	16.4	6855	4 US-09-935-541-3	Sequence 3, Appli
26	51.2	16.4	9839	4 US-09-902-540-996	Sequence 996, App
27	50.8	16.3	2635	3 US-09-126-280-3	Sequence 3, Appli

28	50.8	16.3	2670	3 US-09-126-280-1	Sequence 1, Appli
29	50.8	16.3	2961	2 US-08-407-875-1	Sequence 1, Appli
30	50.8	16.3	2961	3 US-09-277-858-1	Sequence 1, Appli
31	50.4	16.2	1695	4 US-09-902-540-4319	Sequence 4319, Ap
32	50.4	16.2	12730	4 US-09-902-540-1029	Sequence 1029, Ap
33	49.8	16.0	750	4 US-09-252-991A-12577	Sequence 12577, A
34	49.8	16.0	1050	4 US-09-252-991A-12584	Sequence 12584, A
35	49.4	15.8	2589	4 US-09-902-540-4601	Sequence 4601, Ap
36	49.4	15.8	8914	4 US-09-902-540-915	Sequence 915, App
37	49.4	15.8	27490	4 US-09-902-540-1227	Sequence 1227, Ap
38	49	15.7	1341	4 US-09-902-540-5418	Sequence 5418, Ap
39	49	15.7	34953	4 US-09-902-540-1263	Sequence 1263, Ap
40	48.8	15.6	348	4 US-09-270-767-26794	Sequence 26794, A
41	48.8	15.6	962	4 US-09-270-767-11250	Sequence 11250, A
42	48.6	15.6	804	4 US-09-252-991A-4814	Sequence 4814, Ap
43	48.6	15.6	960	4 US-09-252-991A-4705	Sequence 4705, Ap
44	48.6	15.6	1233	4 US-09-252-991A-16145	Sequence 16145, A
45	48.6	15.6	1410	4 US-09-252-991A-16033	Sequence 16033, A

ALIGNMENTS

```
RESULT 1
US-08-765-907A-4
: Sequence 4, Application US/08765907A
: Patent No. 6352839
: GENERAL INFORMATION:
: APPLICANT: BLANC, Veronique
: APPLICANT: THIBAUT, Denis
: APPLICANT: BAMES-JACOUES, Nathalie
: APPLICANT: BLANCHE, Francis
: APPLICANT: COUZET, Joel
: APPLICANT: BARRIBRE, Jean-Claude
: APPLICANT: DEBUSCHE, Laurent
: APPLICANT: FAMECHON, Alain
: APPLICANT: PARIS, Jean-Marc
: APPLICANT: DUTRUC-ROSSERT, Gilles
: TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
: FILE REFERENCE: Streptogramin genes
: CURRENT APPLICATION NUMBER: US/08/765,907A
: CURRENT FILING DATE: 1997-03-20
: NUMBER OF SEQ. ID NOS.: 17
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 4
: LENGTH: 387
: TYPE: DNA
: ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-4
Query Match 43.5%; Score 135.6; DB 3; Length 387;
Best Local Similarity 69.8%; Pred. No. 6.3e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
14 ACCAGCTGCAGCGCGCTGCGCGGAGCTGCAGCCCTTCACGCGGAGCGCTTCTGACACGG 73
59 ACCCCCTGCAGCGCGCTGCGCGGCGCTGCAGCCCGCGCGCGCTGCTGACGCGCG 118
74 TCGCGCGCGCATGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 133
119 TCGGCACACGCTGCGACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 178
134 CGATGATGACGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 193
179 CGATGATGACGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 218
194 ACCAGCGCTGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 253
239 ACCAGCGCATGACGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 298
254 GCGCGCTGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 275
|||||
```

Db 299 GCCGCTCGAGAGCAGTGTGAT 320

RESULT 2
US-09-987-614A-4
Sequence 4, Application US/09987614A
Patent No. 683382
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMBAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSETT, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/09/987,614A
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US/08/765,907
PRIOR FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 4
LENGTH: 387
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-4

Query Match 43.5%; Score 135.6; DB 4; Length 387;
Best Local Similarity 69.8%; Pred. No. 6.3e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGCGCTCTGAGACGG 73
Db 59 ACCCCCTCGACGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTGAGACGGCG 118

QY 74 TCGGCGCGCGCATCGACCTTGGTGTCCGCATCGCGCGGTACAAAGTCCCGGACGGCGTCC 133
Db 119 TCCGCACACCCCTCGACATCTGCTGGCATCGCGCGGTACAAAGTCCCGTCCACAGGTGC 178

QY 134 CGATGATGACCGCGCGCGCGCGCTGCTGTCAGAGACAGAGCGCGCGCGCGCGCGCGCG 193
Db 179 CGATGATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 238

QY 194 ACCACGGCGCTCGACGATCGTCTGGTGAACCTTACGACGTGATCATCAGCGAGATGT 253
Db 239 ACCACGGCGATCGACCGCGCGCTTCTCTGGCGACCTGTACGACGATCATCAGCGAGACT 298

QY 254 GCCGCGTGCAGACCTGTGTAT 275
Db 299 GCCGCTCGAGAGCAGTGTGAT 320

RESULT 3
US-08-765-907A-1
Sequence 1, Application US/08765907A
Patent No. 6352839
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMBAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSETT, Gilles

QY 14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGCGCTCTGAGACGG 73
Db 1931 ACCCCCTCGACGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCTGAGACGGCG 1990

QY 74 TCGGCGCGCGCATCGACCTTGGTGTCCGCATCGCGCGGTACAAAGTCCCGGACGGCGTCC 133
Db 1991 TCCGCACACCCCTCGACATCTGCTGGCATCGCGCGGTACAAAGTCCCGTCCACAGGTGC 2050

QY 134 CGATGATGACCGCGCGCGCGCGGTGTCAGCGCTGTCAGAGACAGAGCGCGCGCGCGCGCG 193
Db 2051 CGATGATGACCGCGCGCGCGCGGTGTCAGCGCTGTCAGAGACAGAGCGCGCGCGCGCG 2110

QY 194 ACCACGGCGCTCGACGATCGTCTGGTGAACCTTACGACGTGATCATCAGCGAGATGT 253
Db 2111 ACCACGGCGATCGACCGCGCGCTTCTGTCGCGACCGCTGTACGACGATCATCAGCGAGACT 2170

QY 254 GCCGCGTGCAGACCTGTGTAT 275
Db 2171 GCCGCTCGAGAGCAGTGTGAT 2192

RESULT 4
US-09-987-614A-1
Sequence 1, Application US/09987614A
Patent No. 683382
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMBAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSETT, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/09/987,614A
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US/08/765,907
PRIOR FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 1
LENGTH: 2888
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Query Match 43.5%; Score 135.6; DB 4; Length 2888;
Best Local Similarity 69.8%; Pred. No. 7.5e-21;
Matches 183; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 14 ACGAGCTGCAGCGCGCTGCGCGGAGCTCGACGCCCTCGACGGAGCGCTCTGAGACGG 73

```
Db      1931 ACCCCCTCGACGGCGTCCGCGCCCGCTGAGACCCCGCGACGCCGCCCTCTGTGAGCCG 1990
      74 TCGCGCGCGCGCATCGACTTCGTGTGTCGCGCATCGCGCGGTACAAGTCCCGCACGGCGTC 133
Db      1991 TCCGCACACGGCTGACATCTGCTGCTGCGATCGCGCAGTCAAGCCGCTCCACCAAGTGC 2050
      134 CGATGATGACGCCCGCGCGGTGAGCTGTGTCAAGACAGAGGCCCGCGCTTACGCCCGCG 193
Db      2051 CGATGATGACGCCCGCGCGGTGAGCTGTGTCAAGACAGAGGCCCGCGCTTACGCCCGCG 2110
      194 ACCAGCGCTCGACGATCGCTTCTGTGTAACCTTACGACGATCATCATCAGGAGTGT 253
Db      2111 ACCAGCGCATCGACCCCGCTTCTGTGCAACCTTGTACAGACGATCATCATCAGGAGTGT 2170
      254 GCGCGCTCGAGAGCTGTGTAT 275
      2171 GCGCGCTCGAGAGCTGTGTAT 2192
```

```
RESULT 5
US-09-252-991A-11493
; Sequence 11493, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11493
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11493
```

```
Query Match      17.5%; Score 54.6; DB 4; Length 507;
Best Local Similarity 49.8%; Pred. No. 0.0029;
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY      33 CGCGAGCTCGACGCCCTCGACCGGACGCTCTGTGACACGCGTCCGCGCGCATCGACT 92
      149 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCCCGCATGCTGTGTGCGCGCGGCTACAT 208
Db      93 CGGTGTCCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
      209 CGCGGTGAGTTCGCTCGATCTTCAACGGGCTCGGCGCGGACACCACTTGTCTACCG 268
Db      153 GGTACACCTGTGTAAAGACAGGCGCGCGCTACGCGCGCGACACAGGCTCGACGAAT 212
      269 TCGGACCTGTTCCTCGCGCGCTTTCGACCGGACGCTCCGAGCATCTGTGCGACGAAT 328
QY      213 GTTCTGTGTGAACCTTACGACGATCATCAGGAGATGTGCGCGTCAAGGACCTGT 272
      329 GGGCAAGAGGCGCTGAGCTGTGAGTTCAACGACGATGCGCCGATGTGACAAAGAGC 388
Db      273 GATGACCGCGGAGAGCTTGAAGCGCGGACGAGACCGGCGG 309
      389 CGATGCGAGCGCTGCGCGGACCTCAAGAGCGGCGG 425
```

```
RESULT 6
US-09-252-991A-11423/c
; Sequence 11423, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
```

```
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11423
; LENGTH: 1359
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11423
```

```
Query Match      17.5%; Score 54.6; DB 4; Length 1359;
Best Local Similarity 49.8%; Pred. No. 0.0031;
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY      33 CGCGAGCTCGACGCCCTCGAGCGGACGCTCTGTGACACGCTGCGCGCGCATCGACT 92
      928 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCGCATGCTGTGTGCGCGCGGCTTACAT 869
Db      93 CGGTGTCCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
      868 CGCGGTGAGTTCGCTCGATCTTCAACGCGCTCGCGCGGACACCACTTGTCTACCG 809
QY      153 GGTACACCTGTGTAAAGACAGGCGCGCGCTACGCGCGCGACACAGGCTCGACGAAT 212
      808 TCGGACCTGTTCCTCGCGCGCTTTCGACCGGACGCTCCGAGCATCTGTGCGACGAAT 749
Db      213 GTTCTGTGTGAACCTTACGACGATCATCAGGAGATGTGCGCGTCAAGGACCTGT 272
      748 GGGCAAGAGGCGCTGAGCTGTGAGTTCAACGACGATGCGCCGATGTGACAAAGAGC 689
QY      273 GATGACCGCGGAGAGCTTGAAGCGCGGACGAGACCGGCGG 309
      688 CGATGCGAGCTGCGCGGACCTCAAGAGCGGCGG 652
Db
```

```
RESULT 7
US-09-252-991A-11470
; Sequence 11470, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11470
; LENGTH: 1371
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11470
```

```
Query Match      17.5%; Score 54.6; DB 4; Length 1371;
Best Local Similarity 49.8%; Pred. No. 0.0031;
Matches 138; Conservative 0; Mismatches 139; Indels 0; Gaps 0;

QY      33 CGCGAGCTCGACGCCCTCGAGCGGACGCTCTGTGACACGCTGCGCGCGCATCGACT 92
      486 CGAAGCGTTCTTCTCTGAGCGCTGCGCGCGCATGCTGTGTGCGCGCGGCTTACAT 545
QY      93 CGGTGTCCGATCGCGCGGTACAAGTCCCGGACGCGCGTCCGATGATGACGCCGCGCG 152
```

Db 546 CGCGCTGAGTTGCTCGATCTTCAACGGGCTCGCGCCGAGACCACTTGTCTACCG 605
Qy 153 GGTGACCTGTGTCAGAGAGAGGCGCGCTACGCGCGGACCAAGGCTCGACGATC 212
Db 606 TGGGACCTGTGTTCTGGGGGCTTTCAGCGGAGCTTCGCGACATCTGCGACCAATT 665
Qy 213 GTTCTGTGTAACTTACGACGTGATCATCAGGAGATGTGCGGCTCGAGGACTTGT 272
Db 666 GGGGCAAGAGGGGCTGACCTGACCTGATTCACAGCGACATCGCCGTATCGACAAACAGGC 725
Qy 273 GATGAGCGGAGAGGCTGACGCGCGGAGCGCGG 309
Db 726 CGATGCGACCTGCGCGGACCTTCAAGAGCGCGG 762

RESULT 8

US-09-902-540-2740
Sequence 2740 Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B
CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 2740
LENGTH: 1963
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-2740

Query Match 17.4%; Score 54.2; DB 4; Length 1963;
Best Local Similarity 51.4%; Pred. No. 0.0039;
Matches 125; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1 ATGACCGAGACAGAGAGCTGACGCGGCTCGCGCGGAGCTGACGCGCTCGACGGAGC 60
Db 1 ATGACCGACCTTCCCGACCTCGACCTCGACCTCGACGAGCGATCGACGAGAG 60
Qy 61 CTCCTGACACGAGTGGGCGCGCATGACCTGAGTGCATGCGGCGGATCAAGTCC 120
Db 61 ATCTGTGATGCTTCCGCGCGGATGCGGCGGATGCGGCGGATGCGGCGGAGCTG 120
Qy 121 CGGACGCGGCTCCGATGATGACCGCGCGGCTGAGCTGATCAAGACAGGCGCGC 180
Db 121 GCGCGCGCGCGCTTCCGCGGACGAGCGCGGAGACCTGCTCTGCGCGCATCCG 180
Qy 181 CGCTAGCGCGCGACGAGCGCTCGACGATGCTTCTGTAACTTCAAGCGATG 240
Db 181 ACCCGCGCGCGAGACGCGCTGACCGCGAGAGTGAAGGACATGCGGCGCTTTC 240
Qy 241 ATC 243
Db 241 ATC 243

RESULT 9

US-09-902-540-1129/c
Sequence 1129 Application US/09902540
Patent No. 6833447

GENERAL INFORMATION:
APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.
TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 1129
LENGTH: 15689
TYPE: DNA
ORGANISM: Myxococcus xanthus
US-09-902-540-1129

Query Match 17.4%; Score 54.2; DB 4; Length 15689;
Best Local Similarity 51.4%; Pred. No. 0.0048;
Matches 125; Conservative 0; Mismatches 118; Indels 0; Gaps 0;

Qy 1 ATGACCGAGACAGAGAGCTGACGCGGCTGCGCGGAGCTGACGCGCTCGACGGAGC 60
Db 8044 ATGACCGACCTTCCCGACCTCGACCTCGACCTCGACCTCGACGAGAG 7985
Qy 61 CTCCTGACACGAGTGGGCGCGCATGACCTGAGTTCGCGATCGCGGATCAAGTCC 120
Db 7984 ATCTGTGATGCTTCCGCGCGGATGCGGCGGATGCGGCGGATGCGGCGGAGCTG 7925
Qy 121 CGGACGCGGCTCCGATGATGACCGCGCGGCTGACGCTTGTCAAGACAGGCGCGC 180
Db 7924 GCGCGCGCGCGCTTCCGCGGACCGCGCGGCTGACGCTTGTCTGCGCGCATCCGC 7865
Qy 181 CGCTAGCGCGCGACGAGCGCTCGACGAGTTCCTGAGTCAAGCGATGATC 240
Db 7864 ACCCGCGCGCGAGACGCGCTGAGCCGCGACGAGTGAAGCGATCTGCGGCTTTC 7805
Qy 241 ATC 243
Db 7804 ATC 7802

RESULT 10

US-08-403-852D-4
Sequence 4 Application US/08403852D
Patent No. 5891695

GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Branche, Francis
APPLICANT: Cruzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
TITLE OF INVENTION: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:

```
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208

US-08-403-852D-4

Query Match      17.3%; Score 54; DB 2; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.0042;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCCCTTCGACGAGGACGCTCCTCGACACGATGCGCGCGCGCATCGACCTCGGTG 97
    |||||
DB 860 AGGTGACCGGTTCGCGCGCGGACGATGCGGTGCGCAAGAACTGTGCGCGCG 919
    |||||

QY 98 TCCGATCGCGCGGTACAAATGCCGACGCGGTCCGATGATGACCGCGCGGTCA 157
    |||||
DB 920 GCTTCGCTCCCGCTGGAAGTCCAGGTGCTGACGATCGGCAAGCGCGCGGTG 979
    |||||

QY 158 GCCTGTCAAGGACGAGCGCGCGCTACGCGCGCGACGCGCGCTGCGAATCGTTC 217
    |||||
DB 980 GCTGTTCGTGACACCTTGCGACCGCGACCGGTGCGCGCGCATGAGAGGCCA 1039
    |||||

QY 218 TGGTGAACCTCTACGAGTATCATCGAGATGTCGCGGTGAGAGACCTGGTATGA 277
    |||||
DB 1040 TCACGAGGTCTTTCGACTGCGCGCGCGCGCATCATCGGACCTGCTGCGGC 1099
    |||||

QY 278 GCCGGAGAGCCTGACGCGCGGAGACCGGC 307
    |||||
DB 1100 CCATCTACGCCGCCACCGCGCTACGCGC 1129
    |||||

RESULT 11
US-08-510-646B-4
; Sequence 4, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Jacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved in The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farbow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
```

```
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208

US-08-510-646B-4

Query Match      17.3%; Score 54; DB 3; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.0042;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCCCTTCGACGAGGACGCTCCTGACACGATGCGCGCGCGCATCGACCTCGGTG 97
    |||||
DB 860 AGGTGACCGGTTCGCGCGCGGACGATGCGGTGCGCAAGAACTGTGCGCGCG 919
    |||||

QY 98 TCCGATCGCGCGGTACAAATGCCGACGCGGTCCGATGATGACCGCGCGGTCA 157
    |||||
DB 920 GCTTCGCTCCCGCTGGAAGTCCAGGTGCTGACGATCGGCAAGCGCGGTG 979
    |||||

QY 158 GCCTGTCAAGGACGAGCGCGCGCTACGCGCGCGACGCGCGCTGCGAATCGTTC 217
    |||||
DB 980 GCTGTTCGTGACACCTTGCGACCGCGACCGGTGCGCGCGCATGAGAGGCCA 1039
    |||||

QY 218 TGGTGAACCTCTACGAGTATCATCGAGATGTCGCGGTGAGAGACCTGGTATGA 277
    |||||
DB 1040 TCACGAGGTCTTTCGACTGCGCGCGCGCGCATCATCGGACCTGCTGCGGC 1099
    |||||

QY 278 GCCGGAGAGCCTGACGCGCGGAGACCGGC 307
    |||||
DB 1100 CCATCTACGCCGCCACCGCGCTACGCGC 1129
    |||||

RESULT 12
US-09-231-818-4
; Sequence 4, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
```

```

1  APPLICANT:  Blanc, Veronique
2  APPLICANT:  Blanche, Francis
3  APPLICANT:  Crouzet, Joel
4  APPLICANT:  Jacques, Nathalie
5  APPLICANT:  Lacroix, Patricia
6  APPLICANT:  Thibaut, Denis
7  APPLICANT:  Zagorec, Monique
8  APPLICANT:  Debussche, Laurent
9  APPLICANT:  De Crecy-Lagard, Valerie
10 TITLE OF INVENTION:  Polypeptides Involved In The
11 TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
12 TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
13 NUMBER OF SEQUENCES:  43
14 CORRESPONDENCE ADDRESS:
15 ADDRESSEE:  Flimegan, Henderson, Farabow, Garrett & Dunner
16 STREET:  1300 I Street, N.W., Suite 700
17 CITY:  Washington
18 STATE:  D.C.
19 COUNTRY:  USA
20 ZIP:  20005-3115
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE:  Floppy disk
23 COMPUTER:  IBM PC compatible
24 OPERATING SYSTEM:  PC-DOS/MS-DOS
25 SOFTWARE:  Patent In Release #1.0, Version #1.30
26 CURRENT APPLICATION DATA:
27 APPLICATION NUMBER:  US/09/231,818
28 FILING DATE:
29 PRIOR APPLICATION DATA:
30 APPLICATION NUMBER:  US/08/403,852
31 FILING DATE:  10-MAY-1995
32 APPLICATION NUMBER:  PCT/FR 93/00923
33 FILING DATE:  25-SEP-1993
34 PRIOR APPLICATION DATA:
35 APPLICATION NUMBER:  FR 92/11441
36 FILING DATE:  25-SEP-1992
37 ATTORNEY/AGENT INFORMATION:
38 NAME:  Meyers, Kenneth J.
39 REGISTRATION NUMBER:  25,146
40 REFERENCE/DOCKET NUMBER:  03806.0054-00000
41 TELECOMMUNICATION INFORMATION:
42 TELEPHONE:  (202) 408-4400
43 TELEFAX:  (202) 408-4400
44 INFORMATION FOR SEQ. ID NO. 4:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH:  1208 base pairs
47 TYPE:  nucleic acid
48 STRANDEDNESS:  double
49 TOPOLOGY:  linear
50 MOLECULE TYPE:  cDNA
51 HYPOTHEICAL:  NO
52 ANTI-SENSE:  NO
53 ORIGINAL SOURCE:
54 ORGANISM:  S.pristinaespiralis
55 FEATURE:
56 NAME/KEY:  CDS
57 LOCATION:  1..1208
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
9
```

Query Match	17.3%	Score 54	DB 3	Length 1208
Best Local Similarity	50.0%	Pred. No. 0.0042		
Matches 135	Conservative 0	Mismatches 135	Indels 0	Gaps 0

Oy	38	AGCTGACAGCCCTTGACGGGACGCTCTCTTGACACGCTGCGCGCGCCGATGACCTCGGTG	97
Db	860	AGGTCGACCCGTTCCGCGCGCGTACCGGATGCGCTGGGTGCGCAAGAAAGTGTGCGCCCGG	919
Oy	98	TCCGATTCGCGCGGTATCAAGTCCGGGACCGGCGTCCGGATGATGCACGCCGGCGGGTCA	157
Db	920	GCTTCGCTCCCGGTGGAGGTCCAGGTGCGCTTACGCCATGCGAAGGCCGAGCGGTGC	979
Oy	158	GCCCTGTCAGAGACAGGCGCGCGCGCTACGCGCGCCGACCAAGGCTTGACGAATCGTTC	217

Db	980	GCCCTGTCGTCGAGACCTTGCGGCACCGGACCGCCGCCAGAGAACCGCATCGAAGAGGCCA	1039
Qy	218	TGCTGGAACCTCTACGACGATATATCAACGAGATGTCCCGCTCGAAGACCTGGTGA	277
Db	1040	TCACCGAGAGCTTTGACCTGCGCGCCCGCGGCATCATCGGACCTGACCTGCTCGCGC	1099
Qy	278	GCCGGGAGAGCTTGACGCGCGGAGAACCGGC	307
Db	1100	CCATCTACGCGCGACCGCGCCGCTTACGGGC	1129

```

RESULT 13
US-09-635-359B-4
: Sequence 4, Application US/09635359B
: Patent No. 6670157
:
: GENERAL INFORMATION:
:
: APPLICANT: Blanc, Veronique
:           Blanche, Francis
:           Crouzet, Joel
:           Jacques, Nathalie
:           Lacroix, Patricia
:           Thibaut, Denis
:           Zagorec, Monique
:           Debussche, Laurent
:           De Crecy-lagard, Valerie
:
: TITLE OF INVENTION: Polypeptides Involved In The
:                       Biosynthesis Of Streptogramins, Nucleotide Sequences
:                       Coding For These Polypeptides And Their Use
:

```

ADDRESSER: Flinnegan, Henderson, Farabow, Garrett & Smith, Inc.
 STREET: 1300 I Street, N.W., Suite 700
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20005-3315
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/635,359B
 FILING DATE: 09-Aug-2000
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 09/231,818
 FILING DATE: 15-JAN-1999
 APPLICATION NUMBER: US 08/403,852
 FILING DATE: 10-MAY-1995
 APPLICATION NUMBER: PCT/FR 93/00923
 FILING DATE: 25-SEP-1993
 APPLICATION NUMBER: FR 92/11441
 FILING DATE: 25-SEP-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Meyers, Kenneth J.
 REGISTRATION NUMBER: 25,146
 REFERENCE/DOCKET NUMBER: 03806.0054-03000
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202) 408-4000
 TELEFAX: (202) 408-4400
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1208 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: S.pristinaespiralis
 FEATURE:
 NAME/KEY: CDS

LOCATION: 1..1208
SEQUENCE DESCRIPTION: SEQ ID NO: 4
US-09-635-359B-4

Query Match 17.3%; Score 54; DB 4; Length 1208;
Best Local Similarity 50.0%; Pred. No. 0.0042;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTGACGCGCTTGACGAGCGCTCTTGACACAGGTGCGCGCCGCGCATGACCTCGGTG 97
DB 860 AGGTGACCGCTTCCGCGCGGACGATGCGCTGGGTGCGCAGAAAGTCTGCTCCGCGG 919
QY 98 TCCGATCGCGCGGTGACAGTCCCGGACGCGGTCCGATGATGACGCGCGCGGATCA 157
DB 920 GCTTCGCTCCCGCTGCGAGGTCCAGTCCGCTTCCGATGCGCAGGCGCGGTGCG 979
QY 158 GCGTGTCAAGAGAGGCGCGCGCTACGCGCGGACGACGCGCTCGACGATCGTTC 217
DB 980 GCGTGTGTCGAGACGTTGCGGACCGGACGCTGCGCCAGAGCGCATGAGAAAGCCA 1039
QY 218 TGGTGAACCTTACGAGTATCATACGAGATGTGCGCGGTGAGAGACTGTGATGA 277
DB 1040 TCACCGAGGTCTTGACCTGCGCGCGCATCATCTCGGACCTGCTGCGGC 1099
QY 278 GCCGGAGAGCGCTGACGCGCGGAGACCGGC 307
DB 1100 CCATCTACGCGCGCCACCGCGCTACGCGC 1129

RESULT 14
US-08-403-852D-1
Sequence 1, Application US/08403852D
Patent No. 5891695
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Branche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/403,852D
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-00000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
US-08-403-852D-1

Query Match 17.3%; Score 54; DB 2; Length 5392;
Best Local Similarity 50.0%; Pred. No. 0.0048;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTGACGCGCTTGACGAGCGCTCTTGACACAGGTGCGCGCCGCGCATGACCTCGGTG 97
DB 4417 AGGTGACCGCTTCCGCGCGGACGATGCGCTGGGTGCGCAGAAAGTCTGCTCCGCGG 4476
QY 98 TCCGATCGCGCGGTGACAGTCCCGGACGCGGTCCGATGATGACGCGCGCGGATCA 157
DB 4477 GCTTCGCTCCCGCTGCGAGGTCCAGTCCGCTTCCGATGCGCAGGCGCGGTGCG 4536
QY 158 GCGTGTCAAGAGAGGCGCGCGCTACGCGCGGACGACGCGCTCGACGATCGTTC 217
DB 4537 GCGTGTGTCGAGACGTTGCGGACCGGACCGCTGCGCCAGAGCGCATGAGAAAGCCA 4596
QY 218 TGGTGAACCTTACGAGTATCATACGAGATGTGCGCGGTGAGAGACTGTGATGA 277
DB 4597 TCACCGAGGTCTTGACCTGCGCGCGCATCATCTCGGACCTGCTGCGGC 4656
QY 278 GCCGGAGAGCGCTGACGCGCGGAGACCGGC 307
DB 4657 CCATCTACGCGCGCCACCGCGCTACGCGC 4686

RESULT 15
US-08-510-646B-1
Sequence 1, Application US/08510646B
Patent No. 6077699
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
APPLICANT: Branche, Francis
APPLICANT: Crouzet, Joel
APPLICANT: Jacques, Nathalie
APPLICANT: Lacroix, Patricia
APPLICANT: Thibaut, Denis
APPLICANT: Zagorec, Monique
APPLICANT: Debussche, Laurent
APPLICANT: De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
TITLE OF INVENTION: Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fimegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/510,646B
FILING DATE: 03-AUG-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03806.0054-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
FAX: (202) 408-4000
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaeapiralis
US-08-510-646R-1

Query Match 17.3%; Score 54; DB 3; Length 5392;
Best Local Similarity 50.0%; Pred. No. 0.0048;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTTCACGCGCCCTCGACGGGACGCTCTCTGACACAGTGCGCGCGCCCGCATGACCTCGGTG 97
DB 4417 AGGTTCACGCGCTTCGCGCGGTACGCGATGCGGTGCGCAAGAAAGTGTGCGCGCG 4476
QY 98 TCCGATCGCGCGGTACAGTCCCGGACCGCGTCCCGATGATGACAGCCCGCGCGGTCA 157
DB 4477 GCTTCGCGCTCCCGTGGAGGTCCAGGTCCCTACCGCATCGCGAGCGCGAGCGGTG 4536
QY 158 GCTGTGTCAGAGACAGGCGCGCGCTACGCGCGCGCAACAGGCTTCGACGAATCGTTCC 217
DB 4537 GCTGTGTCGTCAGAGCGTTCGCGCACCGGACCGTCCCGCGAGCGCATCGAGAGGCCA 4596
QY 218 TGGTGAACCTCTACAGAGTATCATACGAGATGTCCCGCTCGAGAGCTGTGATGA 277
DB 4597 TCACCGAGTCTTCGACCTGCGCGCGCGCATCATCTCGACCTGCTGCGGC 4656
QY 278 GCCGAGAGAGCTGACGCGCGAGAGACCGGC 307
DB 4657 CCATCTACGCGCGCGCGCGCGCTACGCGC 4686

Search completed: October 5, 2005, 10:06:41
Job time : 59.8815 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 : Search time 306.305 Seconds
(without alignments)
7033.027 Million cell updates/sec

Title: US-10-089-514-3

Perfect score: 312
Sequence: 1 atgacgcgagagagacgagc.....cggcgcgagacgcggcgtga 312

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US10F_PUBCOMB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US10G_PUBCOMB.seq:*
- 19: /cgn2_6/ptodata/1/pubpna/US10H_PUBCOMB.seq:*
- 20: /cgn2_6/ptodata/1/pubpna/US10I_PUBCOMB.seq:*
- 21: /cgn2_6/ptodata/1/pubpna/US10J_PUBCOMB.seq:*
- 22: /cgn2_6/ptodata/1/pubpna/US11A_PUBCOMB.seq:*
- 23: /cgn2_6/ptodata/1/pubpna/US11B_PUBCOMB.seq:*
- 24: /cgn2_6/ptodata/1/pubpna/US11C_PUBCOMB.seq:*
- 25: /cgn2_6/ptodata/1/pubpna/US11D_PUBCOMB.seq:*
- 26: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 27: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	312	100.0	312	US-10-472-587-3	Sequence 3, Appl1
2	296.4	95.0	3305	US-10-126-927-68	Sequence 68, Appl1
3	284.8	84.9	12391	US-10-126-927-67	Sequence 67, Appl1
4	59.4	19.0	3114	US-10-437-963-48039	Sequence 48039, A
5	59.2	19.0	838	US-10-425-115-320	Sequence 320, App
6	57.6	18.5	1009	US-10-425-114-33332	Sequence 33332, A
7	55	17.6	1398	US-10-156-761-1302	Sequence 1302, Ap

C	8	55	17.6	9025608	15	US-10-156-761-1	Sequence 1, Appl1
	9	54.8	17.6	1458	17	US-10-107-431-76	Sequence 76, Appl
C	10	54.8	17.6	45055	17	US-10-107-431-277	Sequence 277, Appl
	11	54.8	17.6	9025608	15	US-10-156-761-1	Sequence 1, Appl1
	12	54.4	17.4	1284	15	US-10-156-761-3832	Sequence 3832, Appl
	13	54	17.3	1208	20	US-10-716-803-4	Sequence 4, Appl1
	14	54	17.3	5392	20	US-10-716-803-1	Sequence 1, Appl1
	15	53.6	17.2	837	17	US-10-107-431-112	Sequence 112, Appl
	16	53.6	17.2	2817	15	US-10-156-761-4836	Sequence 4836, Ap
	17	53.6	17.2	29870	21	US-10-494-495-1	Sequence 1, Appl1
C	18	53.6	17.2	45055	17	US-10-107-431-277	Sequence 277, Appl
	19	53	17.0	1356	17	US-10-282-122A-30160	Sequence 30160, A
	20	52.6	16.9	1869	17	US-10-369-493-28376	Sequence 28376, A
	21	52.6	16.9	1869	17	US-10-369-493-31135	Sequence 31135, A
	22	52.6	16.9	2264	19	US-10-437-963-30078	Sequence 30078, A
	23	52.6	16.9	3735	17	US-10-437-963-30078	Sequence 40134, A
	24	52.4	16.8	1410	22	US-10-680-860A-40	Sequence 40, Appl1
	25	52.4	16.8	30943	22	US-10-680-860A-1	Sequence 1, Appl1
	26	52.2	16.7	762	15	US-10-156-761-3402	Sequence 3402, Ap
	27	51.8	16.6	1686	21	US-10-494-495-4	Sequence 4, Appl1
	28	51.6	16.5	732	15	US-10-156-761-1775	Sequence 1775, Ap
	29	51.6	16.5	13654	17	US-10-292-798-1131	Sequence 1131, Ap
	30	51.4	16.5	1393	19	US-10-437-963-8368	Sequence 8368, Ap
	31	51.2	16.4	1707	15	US-10-156-761-4723	Sequence 4723, Ap
	32	51.2	16.4	5562	9	US-09-030-4828-18	Sequence 18, Appl
	33	51.2	16.4	6816	9	US-09-935-541-1	Sequence 1, Appl1
	34	51.2	16.4	6816	16	US-10-425-800-1	Sequence 1, Appl1
	35	51.2	16.4	6855	9	US-09-935-541-3	Sequence 3, Appl1
	36	51.2	16.4	6855	16	US-10-425-800-3	Sequence 8, Appl1
	37	51.2	16.4	6930	19	US-10-377-139-8	Sequence 3, Appl1
	38	50.8	16.3	862	15	US-10-156-761-6124	Sequence 6124, Ap
	39	50.8	16.3	3012	15	US-10-156-761-2716	Sequence 2716, Ap
	40	50.8	16.3	6122	15	US-10-225-567A-179	Sequence 179, Ap
	41	50.8	16.3	6122	19	US-10-755-889-185	Sequence 185, App
	42	50.8	16.3	6122	19	US-10-722-357-23	Sequence 23, Appl
	43	50.6	16.2	1014	15	US-10-156-761-3014	Sequence 3014, Ap
	44	50.6	16.2	1414	20	US-10-739-930-2200	Sequence 2200, Ap
	45	50.6	16.2	1537	20	US-10-425-115-110407	Sequence 110407,

ALIGNMENTS

RESULT 1
US-10-472-587-3
: Sequence 3, Application US/10472587
: Publication No. US20040214274A1
: GENERAL INFORMATION:
: APPLICANT: YANAI, Koji
: APPLICANT: SUMIDA, Naomi
: APPLICANT: WATANABE, Manabu
: APPLICANT: MORIYA, Tatsuaki
: APPLICANT: MORAKAMI, Takehshi
: TITLE OF INVENTION: Transformants Producing Substance P/1022 Derivatives, Methods fo
: FILE REFERENCE: 2003-1302A/MWC/00144
: CURRENT APPLICATION NUMBER: US/10/472,587
: PRIOR FILING DATE: 2003-09-22
: PRIOR APPLICATION NUMBER: 82227/2001
: PRIOR FILING DATE: 2001-03-22
: NUMBER OF SEQ ID NOS: 40
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 3
: LENGTH: 312
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(309)
US-10-472-587-3
Query Match 100.0%; Score 312; DB 20; Length 312;
Best Local Similarity 100.0%; Pred. No. 5.1e-75;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGACCGAGCAAGACGAGCTGACGCGCTGCGCGAGCTCGACGCCCTCGACGCGAGC 60
Db 1 ATGACCGAGCAAGACGAGCTGACGCGCTGCGCGAGCTCGACGCCCTCGACGCGAGC 60

QY 61 CTCTTGACACGGTGGCGCGCGCGATGACCTTGCTGCTCGCATTCGCGCGGTACAAGTCC 120
Db 61 CTCTTGACACGGTGGCGCGCGCGATGACCTTGCTGCTCGCATTCGCGCGGTACAAGTCC 120

QY 121 CGGCACGCGCTCCCGATGATGACGCGCGCGCGCTGACCTTGCTCAAGGACAGCGCGCC 180
Db 121 CGGCACGCGCTCCCGATGATGACGCGCGCGCGCTGACCTTGCTCAAGGACAGCGCGCC 180

QY 181 CGCTACGCGCGCGACCGACGCGCTCGACGAATCGTTCGTGTAACCTCTACGACGTGATC 240
Db 181 CGCTACGCGCGCGACCGACGCGCTCGACGAATCGTTCGTGTAACCTCTACGACGTGATC 240

QY 241 ATGACCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGCCTTGACGCGCGAG 300
Db 241 ATGACCGAGATGTGCGCGCTCGAGGACCTGTGATGAGCGGAGAGCCTTGACGCGCGAG 300

QY 301 GACCGCGCGTGA 312
Db 301 GACCGCGCGTGA 312

RESULT 2

US-10-126-927-68
; Sequence 68, Application US/10126927
; Publication No. US2003082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pastnak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 68
; LENGTH: 3305
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-10-126-927-68

Query Match 95.0%; Score 296.4; DB 14; Length 3305;
Best Local Similarity 98.4%; Pred. No. 7.6e-71;
Matches 311; Conservative 0; Mismatches 1; Indels 4; Gaps 1;

QY 1 ATGACCGAGCAAGACGAGCTGCA----GCGGCTGCGCGCGAGCTTGACGCGCTCTGACGCG 56
Db 2990 ATGACCGAGCAAGACGAGCTGCAAGTTGCGGCTGCGCGAGCTGACGCGCTCTGACGCG 3049

QY 57 GACGCTCTGTCGACACGAGTGGCGCGCGCATGCACTTGAGTCCGCGATCGCGGATACAA 116
Db 3050 GACGCTCTGTCGACACGAGTGGCGCGCGCATGCACTTGAGTCCGCGATCGCGGATACAA 3109

QY 117 GTCCCGGACGCGGTCCGATGATGACGCCCGGCGGATCAGCTGGTCAAGAGACAGGCG 176

Db 3110 GTCCCGGACGCGGTCCGATGATGACGCCCGGCGGATCAGCTGGTCAAGAGACAGGCGC 3169

QY 177 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAATGCTTCTGGTGAACCTCTACAGAGCT 236
Db 3170 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAATGCTTCTGGTGAACCTCTACAGAGCT 3229

QY 237 GATCATCAAGAGATGTGCGCGCTGAGAGACCTGTGATGAGCCGCGAGAGCTTGAAGCGC 296
Db 3230 GATCATCAAGAGATGTGCGCGCTGAGAGACCTGTGATGAGCCGCGAGAGCTTGAAGCGC 3289

QY 297 CGAGGACCGCGGATGA 312
Db 3290 CGAGGACCGCGGATGA 3305

RESULT 3

US-10-126-927-67
; Sequence 67, Application US/10126927
; Publication No. US2003082575A1
; GENERAL INFORMATION:
; APPLICANT: The Scripps Research Institute
; APPLICANT: Schultz, Peter G
; APPLICANT: Wang, Lei
; APPLICANT: Anderson, John C
; APPLICANT: Chin, Jason
; APPLICANT: Liu, David R
; APPLICANT: Magliery, Thomas
; APPLICANT: Meggers, Eric L
; APPLICANT: Mehl, Ryan A
; APPLICANT: Pastnak, Miro
; APPLICANT: Santoro, Stephen W
; APPLICANT: Zhang, Zhiwen
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
; FILE REFERENCE: 54-000120US
; CURRENT APPLICATION NUMBER: US/10/126,927
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/285,030
; PRIOR FILING DATE: 2001-04-19
; PRIOR APPLICATION NUMBER: US 60/355,514
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 79
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 67
; LENGTH: 12391
; TYPE: DNA
; ORGANISM: Plasmid pSCI01, Streptomyces venezuelae papABC
US-10-126-927-67

Query Match 84.9%; Score 264.8; DB 14; Length 12391;
Best Local Similarity 97.9%; Pred. No. 2.6e-62;
Matches 280; Conservative 0; Mismatches 2; Indels 4; Gaps 1;

QY 1 ATGACCGAGCAAGACGAGCTGCA----GCGGCTGCGCGCGAGCTTGACGCGCTCTGACGCG 56
Db 2055 ATGACCGAGCAAGACGAGCTGCAAGTTGCGGCTGCGCGAGCTGACGCGCTCTGACGCG 2114

QY 57 GACGCTCTGTCGACACGAGTGGCGCGCGCATGCACTTGAGTCCGCGATCGCGGATACAA 116
Db 2115 GACGCTCTGTCGACACGAGTGGCGCGCGCATGCACTTGAGTCCGCGATCGCGGATACAA 2174

QY 117 GTCCCGGACGCGGTCCGATGATGACGCCCGGCGGATCAGCTGGTCAAGAGACAGGCGC 176
Db 2175 GTCCCGGACGCGGTCCGATGATGACGCCCGGCGGATCAGCTGGTCAAGAGACAGGCGC 2234

QY 177 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAATGCTTCTGGTGAACCTCTACAGAGCT 236
Db 2235 CGCCGCTGACGCGCGCGACCAAGGCTCGACGAATGCTTCTGGTGAACCTCTACAGAGCT 2294

QY 237 GATCATCAAGAGATGTGCGCGCTGAGAGACCTGTGATGAGCCGCG 282
Db 2295 GATCATCAAGAGATGTGCGCGCTGAGAGACCTGTGATGAGCCGCG 2340


```
RESULT 7
US-10-156-761-1302
; Sequence 1302, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITUKU
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1302
; LENGTH: 1398
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1398)
US-10-156-761-1302
```

```
Query Match          17.6%; Score 55; DB 15; Length 1398;
Best Local Similarity 48.8%; Pred. No. 1.8e-05;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 9 GCAGAACGAGCTGACGCGCTGCGCGGAGCTGACGCCCTGACGGAGCGTCTCTGA 68
    |||||
DB 423 GCAGCACACCCCAACCCCGTACGAGACCGGTCTTCTCTGCGCGCTGCGCGCTGTCGG 482

QY 69 CAGGTCGGCGCGCGCATGCACTCTCGTGTCCGATCGCGGATCAAGTCCCGGACGG 128
    |||||
DB 483 CTTGACCGCGCGGAGCTGCGCGCGCGGCTGCTCGGATCGGCTGCTGCTGCTGCG 542

QY 129 CGTCCGATGATGACGCCCGCGCGGTCAAGCTGTCAGAGAGAGCGCCCGCTACGC 188
    |||||
DB 543 GGTGGCGCTGATGCGCGCGGCACTGCCCGGCTCGCCGCGCAAGAGCGGCTGACCGCGCG 602

QY 189 CGCCGACGAGCGCTGACGAATGTTCTGTGAACCTTACGAGCTGATCAACGA 248
    |||||
DB 603 CGCCGCGCTGTGCTCGCGCGCTCTCCACCCGCTGTGCTCTGATCTGTCGCGGGGCG 662

QY 249 GATGTGCGCGGTGAGAGACCTGTGATGAGCGCGGAGAGCTGACCGGCGAGGACCGGCG 308
    |||||
DB 663 GCACAAAGACCGCTGATGCTGCGGCGCTGCGGCGCGGACTGATGCGCGCGCGCGCG 722

QY 309 GTG 311
    |||
DB 723 GTG 725
```

```
RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHITUKU
; APPLICANT: HATORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
```

```
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (418715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Query Match          17.6%; Score 55; DB 15; Length 9025608;
Best Local Similarity 48.8%; Pred. No. 1.2e-05;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 9 GCAGAACGAGCTGACGCGCTGCGCGGAGCTGACGCCCTGACGGAGCGTCTCTGA 68
    |||||
DB 1624189 GCAGCACACCCCAACCCCGTACGAGACCGGTCTTCTCTGCGCGCTGCGCGCTGTCGG 1624130

QY 69 CAGGTCGGCGCGCGCATGACCTCGGTGTCCGATCGCGGATCAAGTCCCGGACGG 128
    |||||
DB 1624129 CTTGACCGCGCGGAGCTGCGCGCGGCTGCTCGGATCGGCTGCTGCTGCTGCTGCG 1624070

QY 129 CGTCCGATGATGACGCCCGCGCGGTCAAGCTGTCAGAGAGAGCGCCCGCTACGC 188
    |||||
DB 1624069 GGTGGCGCTGATGAGCGCGGCACTGCCCGGCTCGCGCGGCAAGAGGCGTGAACCGCG 1624010

QY 189 CGCCGACGAGCGCTGACGAATGTTCTGTGAACCTTACGAGCTGATCAACGA 248
    |||||
DB 1624009 CGCCGCGCTGTGCTCGCGGCTCTCCACCCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1623950

QY 249 GATGTGCGCGGTGAGAGACCTGTGATGAGCGGAGAGCGGAGCGGAGAGAGCGGCG 308
    |||||
DB 1623949 GCACAAAGACCGCTGATGCTGCGGCGCTGCGCGCGGACTGATGCGCGCGCGCGCG 1623890

QY 309 GTG 311
    |||
DB 1623889 GTG 1623887
```

```
RESULT 9
US-10-107-431-76
; Sequence 76, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 76
; LENGTH: 1458
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-76
```

```
Query Match          17.6%; Score 54.8; DB 17; Length 1458;
Best Local Similarity 49.0%; Pred. No. 2.1e-05;
Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 6 CGAGGAGACGAGCTGACGCGCTGCGCGGAGCTGACGCGCTGACGAGGAGCGTCTCT 65
```

```
Db      684  CCAACGAGACCGCTTCTCGAGAGCCCGCAAGAGACAGACGACTGGAGGGGATGCTCGC 743
        66  GGACACGGGTGGGGCGCCGATCGACTCTGGGTGTCGATCGCGGGTACAAAGTCCCGCA 125
        Db      744  CTGCTGCTCTCAGAGAAACCCGACCTTCGCGACCTCATGGCGGAGGCGATACAGCC 803
        Qy      126  CGGCGTCCCGATGATGAGCCCGGCGGATGAGCTGCTGATCAAGGACAGGGCCCGCTA 185
        Db      804  GGGGCTCCGGGTGAGAGAGACTTCTCTACGTGCGGACAGCTTCCACGGCCCGGCTA 863
        Qy      186  CGCGCGGACGACGCGCTCGACGAATGTTCTGATGAACCTCTACGACGTGATCATCAC 245
        Db      864  CTACCTGCGCGGACGCGCGCTGCTTCCTCGACCCGCTGCTGCACCGGCGTCACT 923
        Qy      246  GGAGATGTGCGCGCTCGAGACCTGCTGATGAGCCGGAGAGCTGACGCGCCGAGAC 303
        Db      924  GGGCTCTACAGCGGATGCTGCGCGGACGCTCGTCTCGCACGCTGAACGAGAC 981
```

RESULT 10

```
US-10-107-431-277/c
; Sequence 277, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Staefle, Alfred
; APPLICANT: Zaopoulou, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTHO
; TITLE OF INVENTION: BIOSYNTHETIC LOCI
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 277
; LENGTH: 45055
; TYPE: DNA
; ORGANISM: Streptomyces mobaraensis
US-10-107-431-277
```

```
Query Match      17.6%; Score 54.8; DB 17; Length 45055;
Best Local Similarity 49.0%; Pred. No. 1.8e-05;
Matches 146; Conservative 0; Mismatches 152; Indels 0; Gaps 0;
```

```
Qy      6  CGACGACGAGCTGACGCGGCTGCGCGGAGACTTCAGCGCTTCAGCGGACGCTCT 65
        Db      37629  CCAACGAGACCGCTTCTCGAGAGCCCGCAAGAGACAGACGACTGAGGCGATGCTCGC 37570
        Qy      66  GGAACGGGTGGCGCGCGCATCGACTCGGTGTCGATCGCGGGTACAAAGTCCCGCA 125
        Db      37569  CTGCTGCTCTCAGAGAAACCCGACCTTCGCGACCTCATGGCGGAGGCGGATACAGCC 37510
        Qy      126  CGGCGTCCCGATGATGAGCCCGGCGGATGAGCTGCTGATCAAGGACAGGGCCCGCTA 185
        Db      37509  GGGGCTCCGGGTGAGAGACTTCTCTACGTGCGGACAGCTTCCACGGCCCGGCTA 37450
        Qy      186  CGCGCGGACGACGCGCTCGACGAATGTTCTGATGAACCTCTACGACGTGATCATCAC 245
        Db      37449  CTACCTGCGCGGACGCGCGCTGCTTCCTCGACCCGCTGCTGCACCGGCGTCACT 37390
        Qy      246  GGAGATGTGCGCGCTCGAGACCTGCTGATGAGCCGGAGAGCTGACGCGCCGAGAC 303
        Db      37389  GGGCTCTACAGCGGATGCTGCGCGGACGCTCGTCTCGCACGCTGAACGAGAC 37332
```

RESULT 11

```
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
```

```
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Query Match      17.6%; Score 54.8; DB 15; Length 9025608;
Best Local Similarity 50.0%; Pred. No. 1.4e-05;
Matches 137; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
```

```
Qy      33  CGGAGACTCGACCGCTCGACCGGACGCTCTGACACAGTGGCGCGCGATCGACT 92
        Db      4753410  CACCGTCTCAAGAGAGTGTGCTGCTCGACGCGGAGTGAAGCGCTACCGTCCG 4753469
        Qy      93  CGGTGTCGATGCGCGGATGACAGTCCGACGCGCTTCGATATGACGCGCGCG 152
        Db      4753470  GCGGCGCTGAGAGAGCGGCTCAAGGTGAGAGACAGCGGCTTCACACGAGGCTCAGCTG 4753529
        Qy      153  GGTACGCTGTGACAGAGACGAGCGCGCTACGCGCGGACCAAGGCTCGACGACATC 212
        Db      4753530  GCGCTTCGAGAGACATAGCGCGGCTGCTCGACTTCCCGCGTGTGACGCGGACGCTC 4753589
        Qy      213  GTTCTGTGAACTCTACGACGTGATCATACGAGATGTGCGGCTGAGAGACTGTG 272
        Db      4753590  CTTCCAGGTGTGTACGAGGAGCGGCTGTGAGCGCGCGGCGGCTTGTGTGATGTGTC 4753649
        Qy      273  GATGAGCGCGGAGAGCTTACGCGCGGACCGG 306
        Db      4753650  GAAAGTCTTGAGAGACAGCGGTACTGAGGCCCG 4753683
```

RESULT 12

```
US-10-156-761-3832
; Sequence 3832, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3832
; LENGTH: 1284
; TYPE: DNA
```

ORGANISM: Streptomyces avermectilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1) .. (1284)
US-10-156-761-3832

Query Match 17.4%; Score 54.4; DB 15; Length 1284;
Best Local Similarity 50.8%; Pred. No. 2,7e-05;
Matches 130; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

33 CGGGAGCTCGACGCGCCCTCGACGGGAGCTCTCGGACACGCGCGCGCATCGACCT 92
1020 CACGAGCTCAGAGAGGTCGTGAGTCTCGACGGGAGAGTAGCGCCCGTACCGCTCG 1079

93 CGGTGTCGCGATCGCGCGGTACAAAGTCCCGGCAAGCGGTCCGATGATGACGCCGCG 152
1080 GCGGGCCCTGAGACGCGCGCTCAAAGTGGAGACAGGGGGGCTTCAACCGAGGCTCAG 1139

153 GGTACGCTGTGTCAGAGACAGGGCGCGCGCTACGCGCGGACACGCGCTCGACGAAATC 212
1140 GCGCTTCGAGGACATGACGCGCGGTGCTGACTTCCCGCGTGGTCAACGCGACGTCA 1199

213 GTTCCTGTGACCTCTACGACGATGATCATCAGGAGATGTCGCGGTGAGAGACCTGCT 272
1200 CTTCCAGGTCTGACGCGAGGCGCGCTGTGGGCGCGCCCGGCTTCGTGATGTGTC 1259

273 GATGACCGCGGAGAGC 288
1260 GAAGGTCTCTGAGAGC 1275

RESULT 13
US-10-716-803-4
Sequence 4, Application US/10716803
Publication No. US20040229236A1
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
Crouzet, Joel
Jacques, Nathalie
Lacroix, Patricia
Thibaut, Denis
Zagorec, Monique
Debreusche, Laurent
De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/716,803
FILING DATE: 20-NOV-2003
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/635,359
FILING DATE: 09-AUG-2000
APPLICATION NUMBER: US 09/231,818
FILING DATE: 15-JAN-1999
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993

APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Arrigo, Salvatore J.
REGISTRATION NUMBER: 46,063
REFERENCE/DOCKET NUMBER: 03806.0054-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1208 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1208
SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-716-803-4

Query Match 17.3%; Score 54; DB 20; Length 1208;
Best Local Similarity 50.0%; Pred. No. 3.4e-05;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

38 AGCTGACGCGCTCGACGGGAGCCTCTGAGACGCTGCGGCGCGCATCGACCTCGGTG 97
860 AGGTGACCGGTTCCGCGCGGTACGCGATCGCTGGGTCCCAAGAACTGTCGCGCGG 919

98 TCCGATCGCGGGGTACAAAGTCCCGGACAGGGGTCCCATGATGACAGCCCGCGGTGA 157
920 GCTTCGCTCCCGCTCGAGGTCGAGGTGCTTACGCTTACGCTTACGCAAGGCGGCGGTG 979

158 GCGTGTCAAGGACAGGGCGCGCGCTACGCGCGGACCGGCTTCGACGATCGTTCC 217
980 GCTTGTCTCGAGACGTTTCGACCGGACCGGCTTGCCTTCCAGAGCGCATCGAAGGCCA 1039

218 TGGTGAACCTTACGACGATCATCAGGAGATGTCGCGCTCGAGGACCTGTGATGA 277
1040 TCACCGAGGCTTTCGACCTGCGCGCGCGGCGCATATCGGACCTCGACCTGTGCGGC 1099

278 GCGGAGAGGCTTACGCGCGCGGAGACCGG 307
1100 CCATCTACGCGCGACCGCGCGCTTACGCGC 1129

RESULT 14
US-10-716-803-1
Sequence 1, Application US/10716803
Publication No. US20040229236A1
GENERAL INFORMATION:
APPLICANT: Blanc, Veronique
Crouzet, Joel
Jacques, Nathalie
Lacroix, Patricia
Thibaut, Denis
Zagorec, Monique
Debreusche, Laurent
De Crecy-Lagard, Valerie
TITLE OF INVENTION: Polypeptides Involved In The
Biosynthesis Of Streptogramins, Nucleotide Sequences
Coding For These Polypeptides And Their Use
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett & Dunner
STREET: 1300 I Street, N.W., Suite 700
CITY: Washington


```
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/716,803
FILING DATE: 20-Nov-2003
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 09/635,359
FILING DATE: 09-AUG-2000
APPLICATION NUMBER: US 09/231,818
FILING DATE: 15-JAN-1999
APPLICATION NUMBER: US 08/403,852
FILING DATE: 10-MAY-1995
APPLICATION NUMBER: PCT/FR 93/00923
FILING DATE: 25-SEP-1993
APPLICATION NUMBER: FR 92/11441
FILING DATE: 25-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Arisgo, Salvatore J.
REGISTRATION NUMBER: 46,063
REFERENCE/DOCKET NUMBER: 03806.0054-04000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 408-4000
TELEFAX: (202) 408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 5392 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: S.pristinaespiralis
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-10-716-803-1

Query Match      17.3%; Score 54; DB 20; Length 5392;
Best Local Similarity 50.0%; Pred. No. 3.2e-05;
Matches 135; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 38 AGCTCGACGCCCTCGACGAGCGCTCTGACACGATGCGCGCGCGCGATCGACCTGGTG 97
DB 4417 AGGTGACCGCTTCGCGCGCGCGCGATGCGCTGGGTGCGCAAGAACGTGCGCGCGG 4476
QY 98 TCCGATCGCGCGGTACATGCCGCGACGCGCGCTCCGATGATGACGCCGCGCGGTCA 157
DB 4477 GCGTCGCGCTCCCGCTGAGAGTCAAGTCCGCTACGCGCATCGGACGCGAGCGGTG 4536
QY 158 GCGTGGTCAAGACAGGCGCGCGCGCTACGCGCGCGCGACGAGCGCTGACGAATGTTCC 217
DB 4537 GCGTGTTCGTGAGACGTTGCGCACCGCGCACCGTCCGCCAGAGCGCATGAGAGGCCA 4596
QY 218 TGTGAACCTCTGACGATGATCATGACGAGATGTGCGCGTGAAGACCTGGTATGA 277
DB 4597 TCACCGAGTCTTTCGACCTGCGCGCGCGCGCGCATTCGCGAGACTTGCTGCGCG 4656
QY 278 GCCGGGAGAGCTGACGCGCGCGAGACCGGC 307
DB 4657 CCATCTACGCGCGCACCGCGCGCTACGCGC 4686
```

```
RESULT 15
US-10-107-431-112
; Sequence 112, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
```

```
APPLICANT: Farnet, Chris
APPLICANT: Staffa, Alfredo
APPLICANT: Zazopoulos, Emmanuel
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND DISTINGUISHING ORTH
FILE REFERENCE: 3001-7US
CURRENT APPLICATION NUMBER: US/10/107,431
CURRENT FILING DATE: 2002-03-28
NUMBER OF SEQ ID NOS: 282
SOFTWARE: Patent version 3.0
SEQ ID NO 112
LENGTH: 837
TYPE: DNA
ORGANISM: Streptomyces mobaraensis
US-10-107-431-112

Query Match      17.2%; Score 53.6; DB 17; Length 837;
Best Local Similarity 52.1%; Pred. No. 4.5e-05;
Matches 147; Conservative 0; Mismatches 129; Indels 6; Gaps 1;

QY 28 CTGCGGCGGAGCTCGACGCGCTCGACGAGAGCTCTCTGACACGCTGCGCGCGCGATC 87
DB 151 CTGCTGACCTCCGCGACGCGCTGCGGAGAGCTCCGCGCGCGCTCCGCGCGCTGAGCTC 210
QY 88 GACCTGGGTGTCGCAATCGCGGCGGTACAGTCCCGGACAGCGGCTCCGATGATGACGCC 147
DB 211 GACTTGCGCGCGCGCGACGTCGCTACCGGCTCTGCTCCGCGCGCGCGAGCTGAGACC 270
QY 148 GCGCGGATGACCTGTGTAAGACAGAGGCGCGCGTACGCGCGCGCGCGCGCGCTGAC 207
DB 271 GCGGATGCGCGCGCGCGCGCGAGACCTGACCGCGCGCGCGCGAGCTGACCGCGCGCTG 330
QY 208 GAATGCTTCTGTGTAACCTTACGACGTGATCATCAGGAGATGTGCGCGCTGAGAGAC 267
DB 331 AGCGCCCTGCGGAGCGGCTGTTCAGCGGAGTCTCTC-----CACCGAGTCTGAG 384
QY 268 CTGGTATGAGCGCGGAGAGCGCTGACGCGCGGAGAGACCGGCGG 309
DB 385 CAGTGACGAGCGCGGACACCGACCTGCGGAGGCGCTCCGCG 426
```

Search completed: October 5, 2005, 09:57:04.
Job time : 331.305 secs

This Page Blank (uspto)

D _b	910 GAGTTTCAGCCGCGAGCTCAGCGACTTATCATACCCGCCGACGTGGCGAGAGGTACGCTCAC	969
Q _y	136 ATGATGCAGCCCCCGCGGTATGCGCTGTCTAAGAACAAGGCGCCCGCTTAGCGCCGAC	195
D _b	970 GTCAAGGCTACGTACAGGTACACCTTATCAGAGCCAACGATCTCTCTCTTGAC	1029
Q _y	196 CACGGCTTCGACGAATCGTTCCTGTGTAACTCTTACGACGATCATCACGAGATGTGC	255
D _b	1030 GTCGGGCTGCGCCAGTACGCCACGACCACCTTCCAAAGTAACGGGTGAACCTGTGGCC	1089
Q _y	256 CGCCTTCAGAACCTGTGTATAGCGCCGGAGAGCTCTGACGCGCCGAGACCGCGC	308
D _b	1090 GGCGGTGTGAAGAAGAGTAAACCCAGAGGAGATTCAGCTGACCGACGGAGCGC	1142

RESULT 2

LOCUS	CN434658	422 bp	mRNA	linear	EST 07-jul-2004
DEFINITION	BE030008B10H1	BE03	Normalized and	Subtracted	bovine embryonic and
			extraembryonic	tissue	Bos taurus cDNA clone BB030008B10H1 5', mRNA
					sequence.
ACCESSION	CN434658				
VERSION	CN434658.1	GI:46413922			
KEYWORDS	EST,				
SOURCE	Bos taurus				
ORGANISM	Bos taurus				(cow)

REFERENCE	1 (bases 1 to 422)
AUTHORS	Lewin, H. A., Renard, J. P., Yang, X. J., Hernandez, A., Degrelle, S.,
TITLE	Hue, I., Tian, X. C., Liu, L. and Everts, R. E.
	Bovine embryonic ESTs

**JOURNAL
COMMENT**

Contact: Harris Lewin
Department of Animal Sciences
University of Illinois at Urbana-Champaign
206 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL 61801

Tel: 217 333 5998
Fax: 217 244 5617
Email: h.lewin@uiuc.edu
Funding for embryonic EST sequencing was provided by USDA ARS contract No. 58-1275-2-020 to H. A. Lewin. Base Calling/Quality Scores: PHRED from Washington University Genome Center. Vector Trimming: cross_match from Washington University Genome Center. PHRAP suite. Sequences submitted are vector free and at least 200 bp in length with average PHRED score > 20. REPEAT IN THE SEQUENCE low_complexity STRAND (+) ELEMENT GC-rich LOCATION [350..383].
PCR Primers

FORWARD: CCGATGACGCGCGTTGTAAACGAC (M13 NIA-F)
 BACKWARD: GTGTGGAATTGGAGGCGATTAACAA (M13 NIA-R)
 Insert Length: 422 Std Error: 0.00
 Plate: BE030008B10 row: H column: 11
 Seq primer: TAATACGACCTCACTATAGGG (T7 PROMOTER)
 High quality sequence count: 422.

FEATURES

Source

```

/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:99131"
/clone="BE030068B10H11"
/tissue_type="embryo (day 36 and day 64) and
extra-embryonic tissue (day 14 to 25) Normalized and
subtracted library"
/lab_host="DH10B"
/clone_id="BE03 Normalized and Subtracted bovine
embryonic and extraembryonic tissue"
name=Vector; GCExfl1(+); Site1: EcoRI; Site2: NotI;
The cDNA library was constructed, normalized and
subtracted by Dr. A. Hernandez, W.M. Keck Center,
University of Illinois Urbana-Champaign and S. Degrelle.

```

ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 422;
Best Local Similarity 51.0%; Pred. No. 0.071;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0

BDR, Institut National de la Recherche Agronomique, France, as described in Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996), *Genome Research* 6(3): 791-806. This clone is from library B503, a normalized library of embryonic tissues substracted with 7800 known placental EST sequences. The double stranded cDNA was size selected (more than 450 bp), adaptorpriced with EcoRI adaptors at both ends and then digested with NotI. The cDNA was then directionally cloned into EcoRI-NotI digested pGEM11zf(+) phagemid vector. Insert size was between 450-2200 bp. Sequencing was done from the 5' end of the clone. Tagged oligo-dT was used to identify the source of the ESTs where possible. (A18)TCGCR : extraembryonic tissue; (A18)TCGCR : 36-day old fetus; (A18)TCGCA = 64-day old fetus (organs); (A18)TCGCA = 64-day old fetus (body) "

[illegible]

RESULT 3

LOCUS	AV610217	653 bp	mRNA	linear	EST 28-NOV-2001
DEFINITION	AV610217 Bos taurus lung fetus Bos taurus CDNA clone E1U031A01 3',				
DESCRIPTION	mRNA sequence.				

VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovinae; Bos.
1 (bases 1 to 653)
Takaue, A., Hirotsune, S., Itoh, R., Jichizono, A., Suzuki, H., Aso, H.

TITLE

JOURNAL Nucleic Acids Res. 29 (22), E108 (2001)
MEDLINE 21570554
PUBMED 11713328
COMMENT Contact: Yoshikazu Sugimoto

Shirakawa Institute of Animal Genetics
Osakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel.: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@cocca.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.


```

High quality sequence stop: 678
FEATURES
Location/Qualifiers
source
1. .678

```

The library was characterised by oligonucleotide fingerprinting (ONFP) to reduce sequencing redundancy. According to the ONFP procedure, clones that display the same hybridisation matrix with a battery of 200 8mer oligonucleotides are grouped into clusters. One clone per ONFP cluster was selected for sequencing. cDNA clones and filters are distributed via Deutsches Ressourcenzentrum fuer Genomforschung GmbH (<http://www.rzpd.de>).

PCR Primers
 FORWARD: 5' CCCAGGCTTACCACTTATCCGGCTCG 3' (M13SP) 5'-seq
 BACKWARD: 5' GCTATTACGCGACGTGCGCAAGGGGATGTG 3' (M13SP) 3'-seq
 Seq primer: 5' GCTATTACGCGACGTGCGCAAGGGGATGTG 3' (M13SP).

FEATURES

source

```

1..712
  /organism="Bos taurus"
  /mol_type="mRNA"
  /db_xref="taxon:9913"
  /clone="R2PD1056J1856Q"
  /sex="female"
  /tissue_type="brain tissue"
  /dev_stage="adult brain"
  /clone_1lb="normal cattle brain"
  /note="Organ: Brain; Vector: pSport1; Site_1: NotI;
  Site_2: SalI; Random primed and directionally cloned in
  pSport1 vector using NotI
  (5'-PGACTAGTCTAGTACGACGCGCGGCC (T)15-3' and SalI 5'-
  TCGACCCAGCGCTCCG-3' adapters (Gibco BRL)."

```

ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 712;
 Best Local Similarity 51.0%; Pred. No. 0.069;
 Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

QY 1 ATACCGAGCAAGAACGCTGCAGCGCTCCGCGGAGCTCGACCGGAGC 60
DB 659 ATATATGCGCGAGACCGTGAAGCGCGCGGTGCTCTTCAACCTTCGACCGTG 600
QY 61 CTCCTGACACGAGTGGCGCGCGCATCGACCTCGGTCCGATCGCGGATCAAGTTC 120
DB 599 CGCGCGCGCATATGATGATGCTGCGGGCGCAAGAGCCGACATGTAACAGGCACC 540
QY 121 CGGACGCGCGTCCCGATGATGACGCCGCGGCTGACGCTGTCAAGACAGGCGCGCC 180
DB 539 GTGAGCTGCTGGCGGAAGATCTCAAGAGAGAGGGGCGAAGCCCTTCTTCAAGGCGCC 480
QY 181 CGTACCGCGCGGACCAAGCGCTTCGACGAACTGCTGTGTAACCTTACGAGTGATC 240
DB 479 TGTCTTCAAGCTTGTGGCGCGCATGGGCGGCGCTGTGCTGCTCTTCAAGAGAGTTC 420
QY 241 ATCAGGAGATGT 253
DB 419 AAGAGGTCTCT 407

```

RESULT 8

CN436828

LOCUS

DEFINITION

714 bp mRNA linear EST 07-JUL-2004
 BE04010B1F12 BE04 Normalized and Subtracted bovine embryonic and
 extraembryonic tissue Bos taurus cDNA clone BE04010B1F12 5', mRNA
 sequence.

ACCESSION CN436828
 VERSION CN436828
 KEYWORDS CN436828.1 GI:46416092

SOURCE EST.
 ORGANISM Bos taurus (cow)

Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovinae; Bos.

REFERENCE 1 (bases 1 to 714)

AUTHORS Lewin,H.A., Renard,J.P., Yang,X.J., Hernandez,A., Degrelle,S.,
 Hue,I., Tiam,X.C., Liu,L. and Everts,R.E.

TITLE Bovine embryonic ESTs
 JOURNAL Unpublished (2004)
 COMMENT Contact: Harris Lewin

Department of Animal Sciences
 University of Illinois at Urbana-Champaign
 206 Edward R. Madigan Laboratory, 1201 W. Gregory Dr., Urbana, IL
 61801, USA

Tel: 217 333 5998
 Fax: 217 244 5617

Email: h-lewin@uiuc.edu

Funding for embryonic EST sequencing was provided by USDA ARS
 contract No. 58-1275-2-020 to H. A. Lewin. Base Calling/Quality
 Scores: PHRED from Washington University Genome Center. Vector
 Trimming: Cross match from Washington University Genome Center
 PHRAP suite. Sequences submitted are vector free and at least 200
 bp in length with average PHRED score > 20. REPEAT IN THE SEQUENCE
 LOW complexity STRAND (+) ELEMENT GC-rich LOCATION [385,418].

PCR PRIMERS
 FORWARD: CCAATCAGCAGCTTGAAGAAGC (M13 NTA-F)
 BACKWARD: GTGTGAATGTGACCGGATACAA (M13 NTA-R)

Insert Length: 714 Std Error: 0.00
 Plate: BE04010B1 row: F column: 12

Seq primer: TAATAGACTCACTATAGG (T7 PROMOTER)
 High quality sequence stop: 714.

FEATURES

source

```

1..714
  /organism="Bos taurus"
  /mol_type="mRNA"
  /db_xref="taxon:9913"
  /clone="BE04010B1F12"
  /tissue_type="embryo (day 36 and day 64) and
  extra-embryonic tissue (day 14 to 25) Normalized and twice
  subtracted library"
  /lab_host="DH10B"
  /clone_1lb="BE04 Normalized and Subtracted bovine
  embryonic and extraembryonic tissue"
  /note="Vector: pGEMZf1(+); Site_1: EcoRI; Site_2: NotI;
  The cDNA library was constructed, normalized and
  substracted by Dr. A. Hernandez, W.M. Keck Center,
  University of Illinois Urbana-Champaign and S. Degrelle,
  INRA, Institut National de la Recherche Agronomique,
  France, as described in Bonaldi, M.F., Lennon, G. and
  Soares, M.B. (1996), Genome Research 6(9): 791-806. This
  clone is from library BE04, a normalized library of
  embryonic tissues substracted with 7800 known placental
  sequences and 5000 clones from the BE03 library. The
  double stranded cDNA was size selected (more than 450 bp),
  adaptored with EcoRI adaptors at both ends and then
  digested with NotI. The cDNA was then directionally cloned
  into EcoRI-NotI digested pGEMZf1(+) phagemid vector.
  Insert size was between 1450-2200 bp. Sequencing was done
  from the 5' end of the clone. Tagged oligo-dT was used to
  identify the source of the ESTs where possible. (A18)TGGCT
  = extraembryonic tissue/ (A18)TGGGT = 36-day old fetus;
  (A18)TGGCA = 64-day old fetus (organs); (A18)TCCGA =
  64-day old fetus (body)"

```

ORIGIN

Query Match 17.5%; Score 54.6; DB 7; Length 714;
 Best Local Similarity 51.0%; Pred. No. 0.069;
 Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

QY 1 ATACCGAGCAAGAACGCTGCAGCGCTCCGCGGAGCTCGACCGGAGC 60
DB 47 ATATATGCGCGAGACCGTGAAGCGCGGTGCTCTTCAACCTTCGACCGTG 106
QY 61 CTCCTGACACGAGTGGCGCGCGCATCGACCTCGGTTCGATCGCGGATCAAGTTC 120
DB 107 CGCGCGCGCATATGATGATGCTGCGGGCGCAAGAGCCGACATGTAACAGGCACC 166
QY 121 CGGACGCGCGTCCCGATGATGACGCCGCGGCTGACGCTGTCAAGACAGGCGCGCC 180
DB 167 GTGAGCTGCTGGCGGAAGATCTCAAGAGAGAGGGGCGAAGCCCTTCTTCAAGGCGCC 226
QY 181 CGTACCGCGCGGACCAAGCGCTTCGACGAACTGCTGTGTAACCTTACGAGTGATC 240

```



```

Db      227 TGCTCAAGCTTTCGGCGGCGATGGCGGCGCCTTGCTGCTGCTCTTACGACGAGCTC 286
QY      241 ATCAGCGAGATGT 253
Db      287 AAGAAGTCATCT 299

RESULT 9
LOCUS   CO882659
DEFINITION Bovgen.10984 normal cattle brain Bos taurus cDNA clone
ACCESSION CO882659
VERSION   R2PDP1056D1724Q 5', mRNA sequence.
KEYWORDS EST.
SOURCE   Bos taurus (cow)
ORGANISM Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovinae; Bos.
REFERENCE 1 (bases 1 to 738)
AUTHORS  Hennig,S., Janitz,M., Herwig,R. and Williams,J.
TITLE    Generation, annotation, evolutionary analysis and database
          integration of 14869 cattle EST clusters
JOURNAL  Unpublished (2004)
COMMENT  Contact: Hennig S
          Laboratory 123, dept. Lehrach
          Max-Planck-Institut fuer Molekulare Genetik
          Ihnestr.63-73, D-14195 Berlin, Germany
          Tel: +49 30 8413 1612
          Fax: +49 30 8413 1380
          Email: hennig@molgen.mpg.de
          The library was characterised by oligonucleotide fingerprinting
          (ONFP) to reduce sequencing redundancy. According to the ONFP
          procedure, clones that display the same hybridisation matrix with a
          battery of 200 8mer oligonucleotides are grouped into clusters. One
          clone per ONFP cluster was selected for sequencing. cDNA clones and
          filters are distributed via Deutsches Ressourcenzentrum fuer
          Genomforschung GmbH (http://www.rzpd.de).
PCR PRIMERS
FORWARD: 5' CCCGAGCTTTACACTTATGCTTCGGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTACGCGACGTCGGGAAGGGGAGATGTG 3' (M13RSP) 3'-seq
Seq primer: 5'-CCGGTCGGAAATCCGGGT-3' (M13RSP).
Location/Qualifiers
1. 738
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /clone="R2PDP1056D1724Q"
   /sex="female"
   /tissue_type="brain tissue"
   /dev_stage="adult brain"
   /clone_lib="normal cattle brain"
   /note="Organ: brain; Vector: pSport1; Site 1: NotI;
   Site 2: SalI; Random primed and directionally cloned in
   pSport1 vector using NotI
   (5'-PCACTGATCTAGATCGGAGCGGCGCGGCC (T)15-3' and SalI 5'-
   TCGACCCACGCGTCCG-3' adapters (Gibco BRL)."

ORIGIN
Query Match      17.5%; Score 54.6; DB 7; Length 738;
Best Local Similarity 51.0%; Pred. No. 0.069;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY      1 ATGACGCGACGAGAGAGCTGACGCGGCTGCGGCGGAGCTGACGCCCTTCAGCGGAGC 60
Db      58 ATGATGCGGAGCGCTGACGCGCGCTGCGGCGGCGGTGCTCTCAACCTTCGACACCGTG 117
QY      61 CTCCTGACACGCGTGGCGCGCGCATGACCTTCGTGTCGCGATCGCGCGGTACAGTCC 120
Db      118 CGGCGCGGATATGATGATGAGTGGGCGCGAAGAGCGGACATCTATGACAAGGGCACC 177
QY      121 CGGACGCGGTCCGATGATGACGCCGCGCGGTGACGCTGTCAAGGACAGGCGCGCC 180

```

```

Db      178 GTGACTGTGCGGAGAGATCTTCAAGGACGAGGCGGCGGAGCCCTTCTTCAAGGCGGCC 237
QY      181 CGCTACGCGCGGACGACGAGCGCTTCGACGAAATGTTCTGTGAACTCTACGAGTGATC 240
Db      238 TGCTCAAGCTTTCGGCGGCGATGGCGGCGCCTTGCTGCTGCTCTTACGACGAGCTC 297
QY      241 ATCAGCGAGATGT 253
Db      298 AAGAAGTCATCT 310

RESULT 10
LOCUS   CR552094
DEFINITION CR552094 Normalized and Subtracted bovine embryonic and
          extraembryonic tissues (bcaz) Bos taurus cDNA clone bcaz0007a.j.09
          5', mRNA sequence.
ACCESSION CR552094
VERSION   CR552094.1 GI:50247004
KEYWORDS EST.
SOURCE   Bos taurus (cow)
ORGANISM Bos taurus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
          Bovinae; Bos.
REFERENCE 1 (bases 1 to 740)
AUTHORS  Renard,J.P., Lewin,H.A., Yang,J., Hernandez,A., Degrelle,S.,
          Everts,R.E. and Hue,I.
TITLE    Embryonic ESTs (bcaz)
JOURNAL  Unpublished (2004)
COMMENT  Contact: Renard JP
          Biologie du Developpement et Reproduction
          INRA
          Domaine de Vilvert 78350 Jouy en Josas, FRANCE
          Email: renard@jouy.inra.fr
          Funding for EST sequencing was provided by funds dedicated to
          J.P. RENARD (BDR, Inra). Funding for construction and subraction of
          the cDNA library was provided by USDA ARS contract No 58-1275-2-020
          to H. A. Lewin The cDNA library was constructed, normalized and
          substracted by Dr. A. Hernandez (W.M. Keck Center, University of
          Illinois Urbana-Champaign) and S. Degrelle (BDR, Inra) as described
          in Bonaldi et al. (1996), Genome Research 6, 791-806. BE03 library
          is a normalized library of embryonic tissues substracted with 7800
          known EST sequences. The double-stranded cDNA was size selected
          (>450 bp), adaptored with EcoRI adaptors at both ends and digested
          with NotI. The cDNA was then directionally cloned into EcoRI-NotI
          digested pEMIZf(+) phagemid vector. Insert size was between
          450-2200 bp. Sequence cleaned of vector, adaptor and repetitions.
          Contact us at sigenasupport@jouy.inra.fr to obtain the chromatogram
          of this sequence.
Location/Qualifiers
1. 740
   /organism="Bos taurus"
   /mol_type="mRNA"
   /db_xref="taxon:9913"
   /clone="bcaz0007a.j.09"
   /tissue_type="embryo (day 36 and day 64), extra-embryonic
   tissue (day 14 to 25)"
   /lab_host="DH10B"
   /clone_lib="Normalized and Subtracted bovine embryonic and
   extraembryonic tissues (bcaz)"
   /note="Vector: pGEMZf1; Tagged oligo-dT was used to
   identify the origin of the ESTs where possible. (A18) TCGGT
   = extraembryonic tissue; (A18) TCGGT = 36-day old fetus;
   (A18) TCGGA = 64-day old fetus (organs); (A18) TCGGA =
   64-day old fetus (body)"

ORIGIN
Query Match      17.5%; Score 54.6; DB 7; Length 740;
Best Local Similarity 51.0%; Pred. No. 0.069;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

```

[illegible]

RESULT 11	LOCUS	DEFINITION
CO874118	745 bp mRNA	linear EST 01-SEP-2004
CO874118	BovGen.02443 normal cattle brain	Bos taurus cDNA clone
R22Dp105601856Q.5'	5', mRNA sequence.	

ACCESSION	CO874118	
VERSION	CO874118.1	GI:5180367
KEYWORDS	EST.	
SOURCE	Bos taurus	(cow)
ORGANISM	Bos taurus	

REFERENCE	1 (bases 1 to 745)
AUTHORS	Hennig, S., Janitz, M., Herwig, R. and Williams, J.
TITLE	Generation, annotation, evolutionary analysis and database integration of 14969 cattle EST clusters unpublished (2004)
JOURNAL	
COMMENT	Contact: Hennig S

laboratory 123, dept. lehrach
Max-Planck-Institut fuer Molekulare Genetik
Inherstr. 63-73, D-14195 Berlin, Germany
Tel.: +49 30 8413 1642
Fax: +49 30 8413 1380
Email: hennigsmoegen.mpg.de
The library was characterised by oligonucleotide fingerprinting
(ONFP) to reduce sequencing redundancy. According to the ONFP
procedure, clones that display the same hybridisation matrix with a
battery of 200 8mer oligonucleotides are grouped into clusters. One
clone per ONFP cluster was selected for sequencing. cDNA clones and
filters are distributed via Deutsches Ressourcenzentrum fuer
Genomforschung GmbH (<http://www.rzpf.de>).

PCR primers
FORWARD: 5' CCCGAGCTTACACTTACATCCCGGCTCG 3' (M13RSP) 5'-seq
BACKWARD: 5' GCTATTTACCCGACGCTGCCAAAGGGAGATGTG 3' (M13FSP) 3'-seq
Seq primer: 5'-CCGAGTCCGAAATCCCGGCT-3' (M13RSP).

FEATURES
SOURCE

```

/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="RZPDp105601856Q"
/sex="female"
/tissue_type="brain tissue"
/dev_stage="adult brain"
/clone_lib="normal cattle brain"
/note="Organ: brain; Vector: pSport1; Site 1: NotI;
Site 2: SalI; Random primed and directionally cloned in
pSport1 vector using NotI
15'-pGACTACTTTCAGATCCGAGACGGGCGCCC (7')15-3' and SalI 5'-
TCGACCCACGCGCTCG-3' adapters (Gibco BRL)"

```

ORIGIN

Query Match	17.5%	Score 54.6;	DB 7;	Length 745;
Best Local Similarity	51.0%;	Pred. No. 0.069;		
Matches 129; Conservative	0;	Mismatches 124;	Indels 0;	Gaps 0;

1 TACCGACAGAAAGACGTGAGGGGCTCGCGCGAGAGCTCGACGGCTCGACGGACG 60
466 ATATATGCGCGAACCCTGACGGCGCTGGGGGGCGTGTCTCTTACCCCTTGACACCGTG 525
61 CTCTTGACACAGATGGCGGCGCGCATCGACTCTGGTGTCCGATCGCGGATCAAGTCC 120
526 CGCGCGCGCATATGATGTGCAGTGGGGGGCAAAAGAGCGCATATTCACAGGGCAC 585
121 CGGACCGCGTCCCGATGATGCAGCCCGCGCGGGTCAAGCTGTGTCAAGACAGGGCGCC 180
586 GTGGACTGTGGCGGAGATCTTCAAGACGAGGGCGGACAGGCCCTTCTTCAAGGGCGCC 645
181 CGGTACGGCGCGACACAGGCGCTCGACAGATGTTCTCGTGAACCTTACAGAGGTATC 240
646 TGGTCCACGTTCTGGCGGACAGGAGCGGCGCTTTCGTGCTGATCTTACGACAGAGCTC 705

241 ATCACGGAGATGT 253
706 AAGAAGTCATCT 718

ESULT 12		/	.
RS53143			
LOCUS			
DEFINITION			
CR551143	754 bp	mRNA	linear
CR551143	Normalized and Subtracted bovine embryonic and		EST 12-JUL-2004

extraembryonic tissues (bcac) Bos taurus cDNA clone bcac0010a.m.23
5', mRNA sequence.
CR553143
CR553143.1 GI:50248053

ETYMOLOGY	EST.
SOURCE	Bos taurus (cow)
ORGANISM	Bos taurus

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 754)
Renard,J.P., Lewin,H.A., Yang,J., Hernandez,A., Degrelle,S.,
Everitts,R.E. and Hue,I.
Embryonic ESTs (bcac)
Unpublished (2004)
Contact: Renard JP

Biologie du Développement et Reproduction
INRA
Domaine de Vilvert 78150 Jouy en Josas, FRANCE
Email: renard@jouy.inra.fr
Funding for EST sequencing was provided by funds dedicated to C.P. RENARD (BDR, Inra). Funding for construction and subtraction of the cDNA library was provided by USDA ARS contract No 58-1275-2-020 to H. A. Lewin. The cDNA library was constructed, normalized and subcloned by Dr. A. Hernandez (W.M. Keck Center, University of Illinois Urbana-Champaign) and S. Degrelle (BDR, Inra) as described in Bonaldo et al. (1996), Genome Research 6, 791-806. B503 library is a normalized library of embryonic tissues subcloned with 7800 known EST sequences. The double-stranded cDNA was size selected (>450 bp), adapted with EcoRI adaptors at both ends and digested with NotI. The cDNA was then directionally cloned into EcoRI-NotI digested pCEMIZ1(+) phagemid vector. Insert size was between 450-2200 bp. Sequence cleaned of vector, adaptor and repetitions. Contact us at siglenasupport@jouy.inra.fr to obtain the chromatogram of this sequence.

FEATURES

source

```

/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="bc20010a.m.23"
/tissue_type="embryo (day 36 and day 64), extra-embryonic
tissue (day 14 to 25)"

```

ORIGIN

ORIGIN

QY	1	ATGACCGGACGAA	CGA	CTCAG	GCGCTG	CGCGGAG	CTTCGA	GCCCTCGA	CGGGACG	60
Db	644	ATGATCGGGACG	ACCGTGA	CGCGCGT	GCGGGCG	CGTGTCTT	CTTCGAC	CCCTTG	CGACACG	585
QY	61	CTCCTGGACA	CGGTGCGG	GCGCGCAT	CGCTCG	GTTCG	CGCATG	CGCGGTA	CAAGTCC	120
Db	584	CGGGGGCGCA	AGATGATG	CAGT	CGGGGCG	CAAAAGG	CGCATCAT	GTATGAA	AGGGCAC	525
QY	121	CGGACCGGCGT	CCCGATGATG	CA	GCGCGCGGG	CGAC	CTGATG	CAAGGA	CAAGGGCGACC	180
Db	524	GTGATCTCTG	TGCGGAGAT	CTCTCA	AGGACGAGGCGG	CAAGGCTT	TTTCA	AGGGCCCC		465
QY	181	CGTACGCGCGG	ACCA	CGGCTCGA	CGAATCGT	TCCTG	GTGAAC	CTTCA	CGACGTGATC	240
Db	464	TGTCCAACG	TTTACGCGG	CATGAGGCGG	CGCTT	CGTGTG	ATCTT	CTTACG	AGAGCTC	405
QY	241	ATCACGAGATCT	253							
Db	404	AAGAGGTCA	TCT 392							

RESULT	15
LOCUS	CR456310
DEFINITION	CR456310 Normalized and Subtracted bovine embryonic and extraembryonic tissues (bcaz) Bos taurus cDNA clone bcaz0018e.n.09
ACCESSION	CR456310
VERSION	CR456310.1
KEYWORDS	GI:49624699
SOURCE	EST.
ORGANISM	Bos taurus (cow)
	Bos taurus
	Emkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovine; Bos
REFERENCE	1 (bases 1 to 812)
AUTHORS	Renard,J.P., Lewin,H.A., Yang,J., Hernandez,A., Degrelle,S., Everts,R.E. and Hue,I.
TITLE	Embryonic ESTs (bcaz)
JOURNAL	Unpublished (2004)
COMMENT	Contact : Renard JP

Domaine de l'ivert 78350 Jouy en Josas, FRANCE
Email: renard@jouy.inra.fr
Funding for EST sequencing was provided by a Genamitil contract from FMS - Ministère de la Recherche (INRA, Num. A020855). Funding for construction and subtraction of the cDNA library was provided by USDA ARS contract No. 58-1225-2-020 to H. A. Lewin. The cDNA library was constructed, normalized and subtracted by Dr A. Hernandez (W.M. Keck Center, University of Illinois (Urbana-Champaign) and S. Degrelle (BDR, Inra) as described in

FEATURES
source

```

ORIGIN
Query Match      17.5%; Score 54.6; DB 7; Length 812;
Best Local Similarity 51.0%; Pred. No. 0.068;
Matches 129; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

    /organism="bos taurus"
    /mol_type="mRNA"
    /db_xref="taxon:9913"
    /clone="bcac0018e.n.09"
    /tissue_type="embryo (day 36 and day 64, extra-embryonic
    tissue (day 14 to 25) "
    /lab_host="DH10B"
    /clone_id="Normalized and Subtracted bovine embryonic and
    extraembryonic tissues (bcac)"
    /note="Vector: pGEMTf1; Tagged oligo-dT was used to
    identify the origin of the ESTs where possible. (A18) TGCCT
    = extraembryonic tissue; (A18) TCGCT = 36-day old fetus;
    (A18) TCGCA = 64-day old fetus (organs); (A18) TCGCA =
    64-day old fetus (body) "

```

ORIGIN

Query March	17.5*	Score 54.6	DB 7	Length 812
Best Local Similarity	51.0*	Pred. No. 0.068		
Matches 123	Conservative 0	Mismatches 124	Indels 0	Gaps 0

Qy	1	ATGACCGAGCAGAACGAGCTGCAGCGGCTGGCGCGAGCTGCAGCGCCTTCGACGGGACG	60
Db	150	ATGATGCGCGAGACCGGTGACGGGCGGTGGCGGGGTGGTCTCTCAACCCCTTCGACACCCTG	209
Qy	61	CTCCTGACACGCGTGGCGGCGCCGCATGCATCTCGGTGCCATTCGCGCGGTACAAGTCC	120
Db	210	CGGCGGGGCATGATGTATGTCAGTGGGGGCGCAAGAGAGCGCATCATGTACAAGGGCAC	269
Qy	121	CGGACGGGCGTCCCCGATGATGACGCCGCGGGGTCACTTGATCAAGACAGGCGCGCC	180
Db	270	GTCGATGCTGCGGGGAGATCTCAAGAGAGAGAGGGCGGCAAGGCGCTTCTTCAAGGGGCGC	329
Qy	181	CGCTACGCGCGCGACCAACGCGCCTCGACGAATCGTTCCTGGTGAACCTTACGACGTATC	240
Db	330	TGGTCCAACTTTCTGGCGGCGCATGGGCGGCGCTTCTGCTGGTCTCTACGACGACCTC	389
Qy	241	ATCACGAGATGT	253
Db	390	AAGAGGTCACTT	402

Search completed: October 5, 2005, 09:01:11
Job time : 1184.18 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 05:31:11 ; Search time 1525.07 Seconds
(without alignments)
3272.568 Million cell updates/sec

Title: US-10-089-514-4
Perfect score: 514
Sequence: 1 MTEQNELQRLAEELDALDGT.....MCRVEDLVMSRESLTAEDRR 103

Scoring table: BLOSUM62
Xgapop 10.0, Ygapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4708233 seqs, 24227607955 residues
Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=xlh
-O=/cgn2_1/USFTO.spool_n/US10089514/runat_04102005_105744_8033/app.query.fasta_1.1621
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514.@CCN_1.1.8782.@runat_04102005_105744_8033 -NCPU=6 -ICPU=3
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hlg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	312	6	BD178314 Transform
2	514	100.0	312	6	BD093915 Streptomy
3	514	100.0	5251	1	AB116234 Streptomy
4	486.5	94.6	14159	1	AF262220 Streptomy

5	256	49.8	387	6	A48325	A48325 Sequence 3
6	256	49.8	387	6	AR198355	AR198355 Sequence
7	256	49.8	2888	6	A48323	A48323 Sequence 1
8	256	49.8	2888	6	AR198353	AR198353 Sequence
9	256	49.8	4740	1	SP060417	SP060417 Streptomy
10	120	23.3	349944	1	BX571871	BX571871 Phototrab
11	120	22.3	349980	6	AX770909	AX770909 Sequence
12	114	22.2	300454	1	AE017175	AE017175 Porphyrom
13	112	21.8	10632	1	U67480	U67480 Methanocald
14	112	21.8	110000	6	AR271569_12	Continuation (13 o
15	109.5	21.3	13242	1	AE000858	AE000858 Methanoba
16	109	20.2	4927	1	AE010450	AE010450 Methanopy
17	107.5	20.9	110000	1	AE017180_19	Continuation (20 o
18	106	20.6	1135	6	AR376132	AR376132 Sequence
19	105	20.4	1105	6	AR387774	AR387774 Sequence
20	104	20.2	21387	1	AE008821	AE008821 Salmonell
21	104	20.2	274050	1	AL627276	AL627276 Salmonell
22	104	20.2	300029	1	AE016842	AE016842 Salmonell
23	103.5	20.1	301630	1	AE016942	AE016942 Bacteroid
24	102	19.8	1152	6	AX457135	AX457135 Sequence
25	102	19.8	3434	1	EHPHEAGN	X60420 Erwinia herb
26	102	19.8	3434	1	ERMP7A	M74133 Erwinia her
27	101	19.6	1253	6	108481	108481 Sequence 3
28	101	19.6	2058	6	108480	108480 Sequence 1
29	101	19.6	3659	6	108487	108487 Sequence 15
30	101	19.6	4190	6	108488	108488 Sequence 19
31	101	19.6	4509	1	ECOPHEAB	M10431 E.coli phe
32	101	19.6	10327	1	AE015281	AE015281 Shigella
33	101	19.6	12071	1	AE005490	AE005490 Escherich
34	101	19.6	16722	1	D90887	D90887 E.coli geno
35	101	19.6	19338	1	D90888	D90888 E.coli geno
36	101	19.6	110000	1	U00096_27	Continuation (28 o
37	101	19.6	270355	1	AP002552	AP002552 Escherich
38	101	19.6	290380	1	AE016987	AE016987 Shigella
39	101	19.6	300099	1	AE016764	AE016764 Escherich
40	100.5	19.6	16905	1	AE005050	AE005050 Halobacte
41	100	19.5	11069	1	AE013693	AE013693 Yersinia
42	100	19.5	110000	1	BX936398_09	Continuation (10 o
43	100	19.5	110000	1	BX936398_10	Continuation (11 o
44	100	19.5	220050	1	AJ414156	AJ414156 Yersinia
45	100	19.5	290803	1	AE017128	AE017128 Yersinia

ALIGNMENTS

RESULT 1	BD178314	312 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD178314				
DEFINITION	Transformant producing p1022 substance and process for producing the same and novel biosynthetic gene.				
ACCESSION	BD178314				
VERSION	BD178314.1	GI:30015579			
KEYWORDS	WO 02077244-A/2.				
SOURCE	Streptomyces venezuelae				
ORGANISM	Streptomyces venezuelae				
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.				
AUTHORS	Yanai,K., Sumida,N., Watanabe,M., Moriya,T. and Murakami,T.				
TITLE	Transformant producing p1022 substance and process for producing the same and novel biosynthetic gene				
JOURNAL	Patent: WO 02077244-A 2 03-OCT-2002; MEIJI SEIKA KAISHA LTD,KOJI YANAI,NAOMI SUMIDA,MANABU WATANABE,TATSUKI MORIYA,TAKESHI MURAKAMI				
COMMENT	OS Streptomyces venezuelae PN WO 02077244-A/2 PD 03-OCT-2002 PF 22-MAR-2002 WO 2002JP002782 PR 22-MAR-2001 JP 01P 082227 PI KOJI YANAI,NAOMI SUMIDA,MANABU WATANABE,TATSUKI MORIYA,TAKESHI MURAKAMI PC C12N15/61,C12N1/15,C12P17/14/(C12N1/15,C12R1:645),(C12P17/14,				

PC C12R1.645)
CC Transformant producing PF1022 substance and process for CC
CC and novel biosynthetic gene
FH Key Location/Qualifiers
FT CDS (1)..(309).
FEATURES
source 1..312
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
ORIGIN
Alignment Scores:
Pred. No.: 9, 73e-47 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-089-514-4 (1-103) x BD178314 (1-312)
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 1 ATGACCGAGCAGAACGAGCTGCAGCGGCTGCGCGGAGCTGCAGCCCTCGACGGGAGC 60
QY 21 LeuLeuAspThrValAlaArgArgAlaLeuAspLeuGlyValArgIleAlaArgTyrLysSer 40
DB 61 CTCCTGGACACAGCTGGGGCGCGCATCGACTCGGTGTCCGATCCGCGGTACCAAGTCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAla 60
DB 121 CGGACGGGGCTCCCATGATGACGCCGCGGCTGACCTGTCAAGACAGGCGGCC 180
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
DB 181 CGCTAGCGCGCGACACAGCGCTCGACGATCGTTCTGTGTGATGAGCGGAGCCTGACGCTGATC 240
QY 81 IleThrGluMetCysArgValAlaGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100
DB 241 ATCACGGAGATGTGGCGCTCGAGGACCTGTGTATGAGCGGAGAGCCTGACGCGCGAG 300
QY 101 AspArgArg 103
DB 301 GACCGGCGG 309
RESULT 2
BD093915 312 bp DNA linear PAT 27-AUG-2002
LOCUS Transformants that produce secondary metabolites modified by a
DEFINITION functional group(s) and novel biosynthesis genes.
ACCESSION BD093915.1 GI:22639503
VERSION WO 0123542-A/2.
KEYWORDS Streptomyces venezuelae
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 312)
Yanai,K., Okakura,K., Yasuda,S., Watanabe,M., Miyamoto,K., Mido,N.
and Murakami,T.
TITLE Transformants that produce secondary metabolites modified by a
functional group(s) and novel biosynthesis genes
JOURNAL Meitai SEIKA KAISHA LTD,KOJI YANAI,KAORU OKAKURA,SHOHEI YASUDA,
MANABU WATANABE,KOICHI MIYAMOTO,NAOKI MIDO,TAKESHI MURAKAMI
OS Streptomyces venezuelae
PN WO 0123542-A/2
PD 05-APR-2001
PF 29-SEP-2000 WO 2000JP006783
PI 29-SEP-1999 JP 99P 276314
COMMENT KOJI YANAI,KAORU OKAKURA,SHOHEI YASUDA,MANABU WATANABE,KOICHI

PI MIYAMOTO,
PI NAOKI MIDO,TAKESHI MURAKAMI
PC C12N15/09,C12N5/10,C12P21/02,C07K11/00// (C12P21/02,C12R1:645)
CC
FH Key Location/Qualifiers
FT CDS (1)..(309).
FEATURES
source 1..312
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/db_xref="taxon:54571"
ORIGIN
Alignment Scores:
Pred. No.: 9, 73e-47 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 6 Gaps: 0
US-10-089-514-4 (1-103) x BD093915 (1-312)
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 1 ATGACCGAGCAGAACGAGCTGCAGCGGCTGCGCGGAGCTGCAGCCCTCGACGGGAGC 60
QY 21 LeuLeuAspThrValAlaArgArgAlaLeuAspLeuGlyValArgIleAlaArgTyrLysSer 40
DB 61 CTCCTGGACACAGCTGGGGCGCGCATCGACTCGGTGTCCGATCCGCGGTACCAAGTCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAla 60
DB 121 CGGACGGGGCTCCCATGATGACGCCGCGGCTGACCTGTCAAGACAGGCGGCC 180
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
DB 181 CGCTAGCGCGCGACACAGCGCTCGACGATCGTTCTGTGTGATGAGCGGAGCCTGACGCTGATC 240
QY 81 IleThrGluMetCysArgValAlaGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100
DB 241 ATCACGGAGATGTGGCGCTCGAGGACCTGTGTATGAGCGGAGAGCCTGACGCGCGAG 300
QY 101 AspArgArg 103
DB 301 GACCGGCGG 309
RESULT 3
AB116234 5251 bp DNA linear BCT 03-JUL-2004
LOCUS Streptomyces venezuelae pApb, pApb, pApb, ORFIV genes for
DEFINITION 4-aminu-4-deoxychorismate synthase, 4-aminu-4-deoxyisoprenate
dehydrogenase, 4-aminu-4-deoxychorismate mutase, hypothetical
protein, complete cds.
ACCESSION AB116234.1 GI:47846865
VERSION AB116234
KEYWORDS Streptomyces venezuelae
SOURCE Streptomyces venezuelae
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1
Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and
Murakami,T.
TITLE Para-position derivatives of fungal antihelminthic cyclopeptides
engineered with Streptomyces venezuelae antibiotic biosynthetic
genes
JOURNAL Nat. Biotechnol. 22 (7), 848-855 (2004)
PUBMED 15184904
REFERENCE 2 (bases 1 to 5251)
AUTHORS Yanai,K., Sumida,N., Okakura,K., Moriya,T., Watanabe,M. and
Murakami,T.
TITLE Direct Submision

JOURNAL

Submitted (01-AUG-2003) Koji Yanai, Meiiji Seika Kaisha, Ltd.,
Microbiological Resources and Technology Laboratories, 788 Kayama,
Odawara-shi, Kanagawa 2500852, Japan
(E-mail:koji_yanai@meiji.co.jp, Tel:81-465-37-5106,
Fax:81-465-37-6397)

FEATURES

Location/Qualifiers

1..5251

SOURCE

/organism="Streptomyces venezuelae"

gene

/mol_type="genomic DNA"

CDS

/db_xref="taxon:54571"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

CDS

/gene="pdaA"

gene

651..2711

ORIGIN

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best local Similarity:

Query Match:

DB:

US-10-089-514-4 (1-103) x AB116234 (1-5251)

QY

1 MetThGluGlnaEngi uleuGlnaArgAlaGluLeuAapAlaLeuAapGlyThr

3673 ATGACGAGCAGAAAGAGCTGACGCGCTGCGCGAGCTCGACGCTCGACGAGCAG

QY

21 LeuLeuAapThrValArgArgArgLeuAapGlyValArgLeuAapGlyThr

3733 CTCTGACACAGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

QY

41 ArgHisGlyValArgProMetMetGlnProGlyArgValSerLeuValysAapAla

3793 CGGACGCGCGCTCCCGATGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

QY

61 ArgTyrAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla

3853 CGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

QY

81 IleThrGluMetCysArgValGluAapLeuValMetSerArgGluSerLeuThrAlaGlu

3913 ATCAGCGAGATGTCGCCCGCTGAGCAGCCTGCTGATGACCGGAGACCTGACGCGCAG

QY

101 AspArgArg 103

3973 GACCGCGCG 3981

Db

RESULT 4

AF262220

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

EVVDLNNAGLVARGAALLVOCGDCAEERFHEAVPDLVRKVDNLQGLAIAIMRAGSGLP
AVAGIAGIAGYGRSSPFTESGAGKPKSYCGAVNEPEEPAPRTPEPRLLTA
YTCOSUVIDEIRPSMGSRGVPLRVRYSHELLLLPYRVRBRAGHYSGAHGWTG
ERTREDGAVHALAQVHNVPYKLPPTUSPEAVALSRLNBEQPGRTFTYRREA
KEVDLEPLPVRAVARHGAIVWLCOPMGNGKLGKHLKRLTEPMKAEYAAVRLTK
EHQOWPAGLHLELTPDPVTECVSELDPRPFTYRSTCDPRLNPEOSADWTFHLSLL

14159 bp DNA linear BCT 09-JAN-2002
Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,
partial sequence.
AF262220
AF262220.2 GI:14290414
Streptomyces venezuelae
Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycinae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 14159)
He,J., Magarvey,N., Pirae,M. and Vining,L.C.
The gene cluster for chloramphenicol biosynthesis in Streptomyces
venezuelae ISP5230 includes novel shikimate pathway homologues and
a monomolecular non-ribosomal peptide synthetase gene
Microbiology 147 (Pt 10), 2817-2829 (2001).
21461106
JOURNAL
MEDLINE
PUBMED
11577160
2 (bases 1 to 14159)
He,J., Magarvey,N.A. and Vining,L.C.
Direct Submission
Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford
St., Halifax, NS B3H 4J1, Canada
3 (bases 1 to 14159)
He,J., Magarvey,N.A. and Vining,L.C.
Direct Submission
Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford
St., Halifax, NS B3H 4J1, Canada

REMARK Sequence update by submitter
COMMENT On Jun 5, 2001 this sequence version replaced gi:10716943.
FEATURES Location/Qualifiers

source

1..14159
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="1S5230"
/db_xref="taxon:54571"
/map="adjacent to pABAB"
2333..3511
/function="essential for chloramphenicol biosynthesis"
/codon_start=1
/transl_table=11
/product="CmlG"
/protein_id="AAG21974.1"
/db_xref="GI:10716944"

CDS

/translation="MTAGSPGAPGTPPLRPALAVRDHISFLADLGMDCDVNCTAFV
GTVEPGRVDEPAMREPVANLGLTRALADIVAEEDVYVTLDPKPGVEGGAELAD
LVTAHNSRPLRIDEAVRLVYMRHIDGSEIVFVFSHLVDLADVDLKQIRAHILG
DTPPGRIGSRYRHARVEQYALPADPVEVTRALGELPVGKGIPTISAVEEML
PLKEGYTFELLSAVTGVLPLAGGGLVLYQYPSRWEFAIKGGYVEIKLVVRPLG
SDYTPYFAKTRHAOSLGRFTWSDLDAPYRFSGRMPVVVSDTFMRPEEDRRM
VPRASARVEEDLKEFLADRSWPGPPLRMQYKRGFLSGAADVLDELKQIRAGSGTP
"

CDS

3498..6404
/note="non-ribosomal peptide synthetase; contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
/product="CmlH"
/protein_id="AAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITDLFRRHRRRTALRTAGRTWYEEELRVTSALARRID
AECPPGRVIVAGEHTAEAVVMAALAMRSVAHTPNNPGLPADRPFEPVAVDAALIV
CEPERALVGEKAGRLAIVAGDVGPDPADPDADDEPARSPVAGSIFTSNGSGID
PKLVYHGGLALNCRSLRLDITPDDOYLTHASISFDASVEIIGTYAGATLVVP
VRQASWLGVSRLAHGCDLMLSPSVARIDEARSKIRVEFCGELGGEYDK
AARYSRVNAVGPTEATVCFSLAELTSTPISIGTPVDGFAVYRDPDSGHATATGE
LVIVGCVNLAGVAGSGPAENVEFGTVDSPAVYTGCVNLSIDGELTYGRIDEQIKR
LGRVNLAVHGSTLSRHLGREVALVRODATILVTAADGEATGESLMARIDLVPWE
APRLVNLALPTLSGGKVDRLRELASPAAPHCCTDGEDAARLRYLDVVTALV
GQIRGPTSTIFDAGSSLMITQYKLSADYGEAVBAAPAMNDYPPAPAPLRHHRG
EAVAPASAVDTLLRRVETBRDLRAELPLRLDTHBEVPGAADREVLTLGASGF
IGHVLDRLLAAGRPLVAVSTGPDGLTGHARFORADYARVALSYAELERWD
RRRGPVDAVHVCQYVNHLLPLDSHLSGCVNTALVVRALAGRSFATSAAG
ADPLPSAATLTAVGDPYRSKLSIEEYVNTLAVGCAVSHYAPGVYGRPEDRHL
KDMVTRTALLETARVGMAPRLSGHVPCDVGTADTLIGRPDANPGTAGASRTPSR
SAVVYRTYALDELRLHGTENDVLAAPAFRRVYDGGGVPPILAAQALSGCGW
PSAHREVDHDLGRLLGTPTDPADRPERTGTTSGPERRRPADTLTGTEP"
3576..3596
/note="Region: adenylation domain A1"
3708..3746
/note="Region: adenylation domain A2"
3939..3986
/note="Region: adenylation domain A3"
4089..4100
/note="Region: adenylation domain A4"
4350..4370
/note="Region: adenylation domain A5"
4494..4538
/note="Region: adenylation domain A6"
4593..4610
/note="Region: adenylation domain A7"
4647..4706
/note="Region: adenylation domain A8"
4833..4853
/note="Region: adenylation domain A9"
4893..4910
/note="Region: adenylation domain A10"
5043..5075
/note="Region: thiolation domain"
5343..5375
/note="Region: NADH binding site"

ORIGIN

Alignment Scores:

Pred. No.: 5,2e-42 Length: 14159
Score: 466.50 Matches: 103
Percent Similarity: 98.108 Conservative: 0
Best Local Similarity: 98.108 Mismatches: 0
Query Match: 94.654 Indels: 2
DB: 1 Gaps: 1

US-10-089-514-4 (1-103) x AF262220 (1-14159)

QY 1 MetThrGluGlnAaGnuLeuGln---ArgLeuArgAlaGluLeuAaPAlaLeuAaPgi 19

Db 11142 ATGACCGAGCAGACAGAGCTGCAGCTTGGCGGTGCCGGGAGCTGAGCGCCTCGACGG 11201

QY 19 YHRLLeuLeuAaPThrValArgArgArg11LeaAaPLeuG1VAlArg11AlaArgTyr1 39

Db 11202 GACGCTTCGAGACCGGTGCGGCGCGCATGACCTGCGTGTCCGATCGCGCGGTACAA 11261

QY 39 sSerATGHiSG1VAlPProMetMetGlnProG1YATGValSerLeuVal1lyAaPArg1 59

Db 11262 GTCCCGGACAGCGGCTGCCATGATGACAGCCCGCGGCTCAGCTGTGTAAGACAGGGC 11321

QY 59 aAlaArgTyrAlaAlaAaPHisG1LeuAaPgiSerPheLeuValAsnLeuTyrAaPva 79

Db 11322 CGCCCGCTACGCGCGCGACCAAGCGCTTCGATGAACTTACCTACGACGT 11381

QY 79 11le1leThrGlnMetCysArgValGlnAaPLeuValMetSerArgGlnSerLeuThra1 99

Db 11382 GATCATACGAGATGTGCGCGTGGAGACCTGTGATGAGCGGGAAGACTGACGGC 11441

QY 99 aGluAaPArgArg 103

Db 11442 CGAGACCGCGCG 11454

Db

11442 CGAGACCGCGCG 11454

RESULT 5

A48325 387 bp DNA linear PAT 07-MAR-1997

LOCUS Sequence 3 from Patent WO9601901.

DEFINITION A48325.1 GI:2302118

VERSION A48325.1 GI:2302118

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

ORIGIN

Alignment Scores:

Pred. No.: 9,32e-19 Length: 387
Score: 256.00 Matches: 51
Percent Similarity: 71.088 Conservative: 8
Best Local Similarity: 61.458 Mismatches: 24
Query Match: 49.814 Indels: 0
DB: 6 Gaps: 0

US-10-089-514-4 (1-103) x A48325 (1-387)


```

source
1..2888
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Alignment Scores:
Pred. NO.: 7,85e-18 Length: 2888
Score: 256.00 Matches: 51
Percent Similarity: 71.08% Conservative: 8
Best Local Similarity: 61.45% Mismatch: 24
Query Match: 49.81% Indels: 0
DB: 6 Gaps: 0

US-10-089-514-4 (1-103) x AR198353 (1-2888)

Oy 7 LeuGlnAArgLeuAArgAlaGlnLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
Db 1936 CTCGACGCGCTCGCGCGCGCGCTCGACCGCGCGCGCTGCTGACGCGCTCGCG 19959
Oy 27 ArgArgGlnLeuAspLeuGlyValArgGlnLeuAlaArgGlyTrpSerArgHisGlyValProMet 46
Db 1996 ACAGCGCTCGACATCGCTCGCTCGCGACGCGCGCGCGCGCGCGCGCTGCGCGCGCGCG 20555
Oy 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgGlyValAlaAspHis 66
Db 2056 ATGACGAGCCCAACCGGATCGCCAGGTCCAGCGCAACGCGCGCGCTGACCGCGCGACAC 21155
Oy 67 GlyLeuAspGlnSerPheLeuValAsnLeuTrpAspValIleIleThrGluMetCysArg 86
Db 2116 GGCATCGACCCCGCTTCCTCGCGACCTGTACGACAGATCATCACCGAGACTGCGCGC 21759
Oy 87 ValGluAsp 89
Db 2176 CTCGAGGAC 2184

RESULT 9
SPU60417 4740 bp DNA linear BCT 07-MAR-1997
LOCUS Streptomyces pristinaeaepristalis 4-dimethylamino-L-phenylalanine
DEFINITION precursor biosynthesis (papa, papC, papB, papM) genes, complete
cde.
ACCESSION U60417
VERSION U60417.1 GI:1575335
SOURCE Streptomyces pristinaeaepristalis
ORGANISM Streptomyces pristinaeaepristalis
KEYWORDS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 4740)
AUTHORS Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Zagorec,M.,
Thiebaut,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and
Thiebaut,D.
TITLE Identification and analysis of genes from Streptomyces
pristinaeaepristalis encoding enzymes involved in the biosynthesis of
the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I
Mol. Microbiol. 23 (2), 191-202 (1997)
JOURNAL MEDLINE
PUBMED 9044253
REFERENCES 2 (bases 1 to 4740)
AUTHORS Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Thiebaut,J.,
Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and Thiebaut,D.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc
Roter, 13 quai Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403,
France

```

gene	gene	gene	gene
<p>complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/function="dehydrogenase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PAPC"</p> <p>/protein_id="AAC44867.1"</p> <p>/db_xref="GI:1575337"</p> <p>/translation="MRGGSVFGRCVVYGGAGAVGRMFSHMLVBSGVAVTMLDVAGAGAADCGRVYAGDVRRRGPEPAVAALAAADVLTVAVEPVAMEVEVLACGMRCAYLADTLTGVSRVLAARLEAPGLOAVGLNPMFNPESLGLQRPVAAVVTDGSEVRALVELVAMGAVEMPAARLRHDELTAQQAATHAAVLAALGELISVDVGLRSDGAPPHLMALLLACGTGPEVYFDIQANAGAPAAQAOLGRGVRLTQGAVERGDEETFAALFELRGVLGEGHAEIRLRCARMTALH"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>/function="mutase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabB"</p> <p>/protein_id="AAC44868.1"</p> <p>/db_xref="GI:1575338"</p> <p>/translation="MTPPAPLPAAPATGPAATDPDLARLLADDAADLLDVRTRLDICRIGEYKRLHOVPMOPHRIAOVNAANARVAADHGIDPAFLRTIYDTLIITETCLLEDRIASGCAVPPTPVHASASARGAVS"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>/function="N-methylase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabM"</p> <p>/protein_id="AAC44869.1"</p> <p>/db_xref="GI:1575339"</p> <p>/translation="MTAATATTLAQAALDEATGOLTGAGITADARADTRLAAACQVAPGDLDTCLAGVPVPRFMHYRRRLTREPARITGHAFTMKHARDLADGCVVPRKETTEEITDADARLEALVRGTTAPLVLDLGGPSTYMAVTLAHVPAARVIGITELISQAARAARRAGTGARIVQGDADAPBELSGDVLVTNPPIPIGLRTSAPBEVLHDPPLALMAGCEGLGMIRAMERTARLLAPGCVLLLEHGSYOLASVPALFPATGWSHASSRPTCNDCGLTAARNHTCAPPA"</p>	<p>complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/function="dehydrogenase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PAPC"</p> <p>/protein_id="AAC44867.1"</p> <p>/db_xref="GI:1575337"</p> <p>/translation="MRGGSVFGRCVVYGGAGAVGRMFSHMLVBSGVAVTMLDVAGAGAADCGRVYAGDVRRRGPEPAVAALAAADVLTVAVEPVAMEVEVLACGMRCAYLADTLTGVSRVLAARLEAPGLOAVGLNPMFNPESLGLQRPVAAVVTDGSEVRALVELVAMGAVEMPAARLRHDELTAQQAATHAAVLAALGELISVDVGLRSDGAPPHLMALLLACGTGPEVYFDIQANAGAPAAQAOLGRGVRLTQGAVERGDEETFAALFELRGVLGEGHAEIRLRCARMTALH"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>/function="mutase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabB"</p> <p>/protein_id="AAC44868.1"</p> <p>/db_xref="GI:1575338"</p> <p>/translation="MTPPAPLPAAPATGPAATDPDLARLLADDAADLLDVRTRLDICRIGEYKRLHOVPMOPHRIAOVNAANARVAADHGIDPAFLRTIYDTLIITETCLLEDRIASGCAVPPTPVHASASARGAVS"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>/function="N-methylase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabM"</p> <p>/protein_id="AAC44869.1"</p> <p>/db_xref="GI:1575339"</p> <p>/translation="MTAATATTLAQAALDEATGOLTGAGITADARADTRLAAACQVAPGDLDTCLAGVPVPRFMHYRRRLTREPARITGHAFTMKHARDLADGCVVPRKETTEEITDADARLEALVRGTTAPLVLDLGGPSTYMAVTLAHVPAARVIGITELISQAARAARRAGTGARIVQGDADAPBELSGDVLVTNPPIPIGLRTSAPBEVLHDPPLALMAGCEGLGMIRAMERTARLLAPGCVLLLEHGSYOLASVPALFPATGWSHASSRPTCNDCGLTAARNHTCAPPA"</p>	<p>complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/function="dehydrogenase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PAPC"</p> <p>/protein_id="AAC44867.1"</p> <p>/db_xref="GI:1575337"</p> <p>/translation="MRGGSVFGRCVVYGGAGAVGRMFSHMLVBSGVAVTMLDVAGAGAADCGRVYAGDVRRRGPEPAVAALAAADVLTVAVEPVAMEVEVLACGMRCAYLADTLTGVSRVLAARLEAPGLOAVGLNPMFNPESLGLQRPVAAVVTDGSEVRALVELVAMGAVEMPAARLRHDELTAQQAATHAAVLAALGELISVDVGLRSDGAPPHLMALLLACGTGPEVYFDIQANAGAPAAQAOLGRGVRLTQGAVERGDEETFAALFELRGVLGEGHAEIRLRCARMTALH"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>/function="mutase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabB"</p> <p>/protein_id="AAC44868.1"</p> <p>/db_xref="GI:1575338"</p> <p>/translation="MTPPAPLPAAPATGPAATDPDLARLLADDAADLLDVRTRLDICRIGEYKRLHOVPMOPHRIAOVNAANARVAADHGIDPAFLRTIYDTLIITETCLLEDRIASGCAVPPTPVHASASARGAVS"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>/function="N-methylase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabM"</p> <p>/protein_id="AAC44869.1"</p> <p>/db_xref="GI:1575339"</p> <p>/translation="MTAATATTLAQAALDEATGOLTGAGITADARADTRLAAACQVAPGDLDTCLAGVPVPRFMHYRRRLTREPARITGHAFTMKHARDLADGCVVPRKETTEEITDADARLEALVRGTTAPLVLDLGGPSTYMAVTLAHVPAARVIGITELISQAARAARRAGTGARIVQGDADAPBELSGDVLVTNPPIPIGLRTSAPBEVLHDPPLALMAGCEGLGMIRAMERTARLLAPGCVLLLEHGSYOLASVPALFPATGWSHASSRPTCNDCGLTAARNHTCAPPA"</p>	<p>complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/complement (2486 .3376)</p> <p>/gene="papC"</p> <p>/function="dehydrogenase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PAPC"</p> <p>/protein_id="AAC44867.1"</p> <p>/db_xref="GI:1575337"</p> <p>/translation="MRGGSVFGRCVVYGGAGAVGRMFSHMLVBSGVAVTMLDVAGAGAADCGRVYAGDVRRRGPEPAVAALAAADVLTVAVEPVAMEVEVLACGMRCAYLADTLTGVSRVLAARLEAPGLOAVGLNPMFNPESLGLQRPVAAVVTDGSEVRALVELVAMGAVEMPAARLRHDELTAQQAATHAAVLAALGELISVDVGLRSDGAPPHLMALLLACGTGPEVYFDIQANAGAPAAQAOLGRGVRLTQGAVERGDEETFAALFELRGVLGEGHAEIRLRCARMTALH"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>3413 .3802</p> <p>/gene="pabB"</p> <p>/function="mutase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabB"</p> <p>/protein_id="AAC44868.1"</p> <p>/db_xref="GI:1575338"</p> <p>/translation="MTPPAPLPAAPATGPAATDPDLARLLADDAADLLDVRTRLDICRIGEYKRLHOVPMOPHRIAOVNAANARVAADHGIDPAFLRTIYDTLIITETCLLEDRIASGCAVPPTPVHASASARGAVS"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>3799 .4677</p> <p>/gene="pabM"</p> <p>/function="N-methylase"</p> <p>/codon_start=1</p> <p>/transl_table=1</p> <p>/product="PabM"</p> <p>/protein_id="AAC44869.1"</p> <p>/db_xref="GI:1575339"</p> <p>/translation="MTAATATTLAQAALDEATGOLTGAGITADARADTRLAAACQVAPGDLDTCLAGVPVPRFMHYRRRLTREPARITGHAFTMKHARDLADGCVVPRKETTEEITDADARLEALVRGTTAPLVLDLGGPSTYMAVTLAHVPAARVIGITELISQAARAARRAGTGARIVQGDADAPBELSGDVLVTNPPIPIGLRTSAPBEVLHDPPLALMAGCEGLGMIRAMERTARLLAPGCVLLLEHGSYOLASVPALFPATGWSHASSRPTCNDCGLTAARNHTCAPPA"</p>

DB:	1	Gaps:	0	
US-10-089-514-4 (1-103) × SP060417 (1-4740)				
OY	7	LeuGlnA ₁ gLeuArgA ₁ gLeuLeuAspA ₁ leuAspGlyThrLeuLeuAspThrValArg	26	
Db	3476	CTTCAGCGGGCTGCGGGCCCGCTGAGACCCCGGAGCGCGCCCTGCTGAGACCGCTCCG	3535	
OY	27	ArgArgIleAspLeuGlyValArgIleAlaArgTyrTyrSerArgHisGlyValProMet	46	
Db	3536	ACACGCCCTGACATCTGCCTGCGCATCGGACGATCAAGACCGCTCCACCAAGTGGCCGATG	3595	
OY	47	MetGlnP ₁ ProGlyArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAspHis	66	
Db	3596	ATGACAGCCCAACCCGATGCCACGATGCCACGCCAACGCCGCCGCTAGCGCCGCGACAC	3655	
OY	67	GlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIleIleThrGlnMetCysArg	86	
Db	3656	GGCATGACCCCGCTCTCTGCGCACCTTGACACAGATATATACCGAAGCTGCGCC	3715	
OY	87	ValGlnAsp	89	
Db	3716	CTCAGGAGAC	3724	
RESULT 10				
LOCUS	BX571871	349944 bp	DNA linear BCT 26-SEP-2003	
DEFINITION	Photorhabdus luminescens subsp. laumondii T101 complete genome;			
ACCESSION	BX571871	BX470251		
VERSION	BX571871.1	GI:36786846		
KEYWORDS	complete genome.			
SOURCE	Photorhabdus luminescens subsp. laumondii T101			
ORGANISM	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.			
REFERENCE	1	Duchaud, E., Rusniok, C., Frangeul, L., Buchrieser, C., Taourit, S., Bocs, S., Bouraux-Eude, C., Chandlier, M., Dassa, E., Derose, R., Derzelle, S., Freysinet, G., Gaudreau, S., Glavaud, A., Glaeser, P., Meizelle, C., Lanos, A., Powell, K., Siquier, P., Wingate, V., Zouine, M., Bomare, N., Danchin, A. and Kunst, F. Complete genome sequence of the entomopathogenic bacterium Photorhabdus luminescens Nat. Biotechnol. 11 (1) (2003) In press		
AUTHORS	2	Duchaud, E., Frangeul, L., Rusniok, C. and Kunst, F. Direct Submission Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: lfrangeul@pasteur.fr, fkunst@pasteur.fr		
JOURNAL		location/Qualifiers		
TITLE		1. 349944		
JOURNAL		/organism="Photorhabdus luminescens subsp. laumondii T101"		
FEATURES		/mol_type="genomic DNA"		
SOURCE		/strain="T101"		
		/db_xref="taxon:243265"		
		complement(1142..3038)		
		/gene="prpB"		
		/locus_tag="plu3539"		
		complement(1142..3028)		
		/gene="prpE"		
		/locus_tag="plu3539"		
		/codon_start=1		
		/transl_table=11		
		/product="Propionate--CoA ligase (Propionyl-CoA synthetase)"		
		/protein_id="CAE35912.1"		
		/db_xref="GI:36786847"		
		/translation="MSFAFYQSGIDPNAMFAAOAORIPWQHPFEVDLYDSNPPARWFGCCATCNALDPMLESDPADKALITSTNSKRVFPRKILHGVNRAAMKSLGKKGGDRVLYYPMIALBALVYLACIAGVHVSIVGCFASGSLATRLDNEPVLVVSADGARGCKIIPYKPLIDEAILELANHPRVLVNVNGLMVMVPPGDDIDFATLRCK		

[illegible]

RBS
INYYDFEOKLIDFLFOKSO"
complement (6731.. 6736)
/gene="prpb"
/locus_tag="plu3542"
7047.. 8724
/gene="prpr"
/locus_tag="plu3543"
7047.. 7052
/gene="prpr"
/locus_tag="plu3543"
7057.. 8724
/gene="prpr"
/locus_tag="plu3543"
/codon_start=1
/transl_table=1
/product="Propionate catabolism operon regulatory protein"
/protein_id="CAE15916.1"
/db_xref="GI:36786851"
/translation="MAETPSAERDNRDKPVIMTWVSRLPDLPRDIPERFDHONITP
IOLGERMVCHIRORLTTECCDNYLCAGSNAYAKSLSPVITAKASGDFMOLAR
ARQISRGVITVQNTLPALERFOORFMRLEORSYTEEDARQVNAKAIIGIKTV
GAGLTIDAYEAEILGVFVSADSIROMFHALDIAWTTLSQASISAYPTENNLRV
HTINDPDSAVMEHVRNTIMLYARSPATVLIQESGKELVQALIHRYALRYGV
AGKHORPVANCAIABLLEALFGEBCAFPGSGRRGAGLEVAHHGTLELFI
GEMPLPLOTLLRYLVEEKSVVRVGGHRIANDVRIICATHCDDLTWODRPFYDIP
RLSVARINIPALREKGDIGILAEYLRARPAALNLYSOSRVELATCOBALDHYH
PENVEKLEMERLTLVYSAHEHILPELILASPERSQSLKQAEQSMKSPSSIQQ
SLVLPVLSHETLKRFGSGNHHAITYLGISRTYIMRLKKEIAREE"
complement (8738.. 9709)
/locus_tag="plu3544"
complement (8738.. 9709)
/locus_tag="plu3544"
/function="Unknown"
/notes="unnamed protein product; similar to probable
transcription regulator"
/codon_start=1
/transl_table=1
/protein_id="CAE15917.1"
/db_xref="GI:36786852"
/translation="MNITKILFLFPNVNLDLSPLQVPTTANELAEKGRHRRPVAI
KVADLPQWTOAGITVQALPEPLDSDPADTLIVAGGIVGESANHPPLRNMLKGA
SLARIVVSGCAFLIACGLDGRVYVTHMSGCAELARYSKLRVCSDPIFDQDVG
WTSAGVTGIDILALALEEDRGHKLADVADYIMFLKRGCGQCFAPLALDSSTSK
PADLHAMTNNIAGDLSVATITATPANSERSFVTHYCHAGITPAKVERFRLEAARH
LVSEITLKLKICACGFCGCEDTLRRSFRAFSSTPDYRKHFKTTP"
10003.. 10784
/locus_tag="plu3545"
10003.. 10008
/locus_tag="plu3545"
10017.. 10784
/locus_tag="plu3545"
/function="Unknown"
/notes="unnamed protein product; Some similarities with
unknown protein"
/codon_start=1
/transl_table=1
/protein_id="CAE15918.1"
/db_xref="GI:36786853"
/translation="MKNTSKHPALASLIINWASGEKPTLAFFVIFGMPFLMDIMGAT
VLGSGMNITVYWRDLIPRTEHLAHISFSQSTPRFQONKVDICVGTGCTNPALLAE
IDMLDHIKVGSKASWTSVCTGSLILGAGLGLGQALAHWSNIDUNKFDPIVVK
KVIDIDRNVITGGGVTAGIDFGLTLIAELKGEVAVQVLAQMYSPEPPNSGMPEDAF
KEITATVQAIFSNVEKQTPYKEILIKSAKRISDIY"
complement (1153.. 12540)
/gene="ISPlu15A"
/locus_tag="plu3546"
complement (11533.. 12540)
/gene="ISPlu15A"
/locus_tag="plu3546"
/function="Transposase"
/codon_start=1
/transl_table=1
/product="Transposase, ISNCY family"

US-10-089-514-4 (1-103) x BX571871 (1-349944)		
Alignment Scores:	Score: 0.632	Length: 349944
Pred. No.:	Score: 120.00	Matches: 25
Percent Similarity:	55.06%	Conservative: 24
Best Local Similarity:	28.09%	Mismatches: 40
Query Match:	23.35%	Indels: 0
DB:	1	Gaps: 0
<pre> /!protein_id="CAE15919.1" /db_xref="GI:36786854" /translation="MKKNTPTTHDAIFPKFUSHITPADPFLIEHLPAFLRAYCDLDT LRLESGSFIEDNLRAVHYSDILYSKTPQESVYVCVIEHQSSDPDKMARLIMKYSISA MOMHEGKHEKEDLRLVLPVLFPGKIRPYPWSTWMEPCFPASALAEIYSAPLAVDT VLPDEELTHKRVALLIEVOKHIFRORDMELQDELTMFLPAVDYVTEILKSMINTYLL VGP7ADPEGFIQLAQQPFXEVEVMTINQIKHKGHSGIKGGLQKQDAREGLOE </pre>		
Oy	1	1 Metchrclucflnansnglulenglnaaglyleuargalagluleuapralaleuaprglythr 20
Db	32716	ATTAATATAACAGGAATATTACACCTTACAGGGAATCTTTGGATGTTATTAAATATGCAG 32655
Oy	21	Leuenuapshryalargargargileapleapglyvalargilealarglyrlyser 40
Db	32856	ATTTTGGCACTGCTCTTGTGAACCTATGAAGAATTGTATGAATAATCCCTGAGGTAAGGCG 32599
Oy	41	Arghtisglyvalprowetmetginproglyargvalserleuvallysapargalala 60
Db	32596	GGGAGGATATTCGATGATGACGCCACACGATATACCTCGTTATGATATGTTGCCG 32533
Oy	61	Arglyrvalalaleaphisglyleuaprgluserphleuvalasleuvalservalle 80
Db	32536	GATTAATCAACGATTTTGGSTTACGCGCTCGAATATATCGAATCAATATTCAACGTGT 32477
Oy	81	l!ethrglmetcysargvalgluap 89
Db	32476	ATTGAAGAGACTTGCCCTCGAAGAGAA 32450
RESULT 11		
LOCUS	AX770909	349980 bp DNA linear PAT 02-JUL-2003
DEFINITION	Sequence 40 from Patent WO02094867.	
ACCESSION	AX770909	
VERSION	AX770909.1	GI:32438073
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS	1	Photorhabdus luminescens
TITLE	Duchaud, E., Taouil, S., Glaeser, P., Frangeul, L., Kunst, F., Danchin, A. and Buchrieser, C.	Photorhabdus luminescens
JOURNAL	Sequence of the photorhabdus luminescens strain T101 genome and uses	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; Photorhabdus.
FEATURES	Patent: WO 02094867-A 40 28-NOV-2002; INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (CNRS) (FR)	
SOURCE	Location/Qualifiers	
	1. 349980	
	/organism="Photorhabdus luminescens"	
	/mol_type="genomic DNA"	
	/db_xref="taxon:29488"	
	/note="sequence length too big . Splittd into 2 sequences. ~seq 00040: 1 to 349980-seq 10790: 300001 to 649980-seq 10791: 600001 to 666798"	
ORIGIN		
Alignment Scores:	Score: 0.632	Length: 349980
Pred. No.:	Score: 120.00	Matches: 25
Percent Similarity:	55.06%	Conservative: 24
Best Local Similarity:	28.09%	Mismatches: 40
Query Match:	23.35%	Indels: 0

DB:	6	Gaps:	0
US-10-089-514-4 (1-103) x AX770909 (1-349980)			
OY	1	MethTrgLUgLaSncgluLeuGlnArgLeuArgLaGluLeuSepalaLeuAepgLYthr 20	
Db	113584	ATTAATTAACACGAAATATTACCACTTACAGGGAATCTTTGATGTATTATTAATGACAG 113643	
OY	21	LeuLeuAepThrValArgArgArgArgLeuAepLeuGluValArgArgLeuAepgLYthr 40	
Db	113644	ATTTGGCAGCTGCTTCTTCAAGCTATGAAAGATTGTATGAAATTCCTGACGTAAAGCG 113703	
OY	41	ArgHisgLYValPProMetMetGlnProGlyArgValSerLeuValLysSepaArgala 60	
Db	113704	GGCGAGATATTCGATGATGACGCCACACGATTAATCTCGTATGATTAATGTTGGCG 113763	
OY	61	ArgTrValaAlaAspHisgLYLeuAepgLYSerPheLeuValaAsnLeuTrpAepVal 80	
Db	113764	GATTAATCAACCGATTTGGGTTACGCGCTGATATATATGATCAATCAATTTCACTGGTT 113823	
OY	81	LeuTrgLUmetCysArgValgLuSep 89	
Db	113824	ATTGAAGAGACTTGCCGCTGAGAGAA 113850	
RESULT 12			
LOCUS	AE017175	300454 bp	DNA linear BCT 02-SEP-2003
DEFINITION	Porphyromonas gingivalis W83 section 4 of 8 of the complete genome.		
ACCESSION	AE017175	AE015924	
VERSION	AE017175.1	GI:34396931	
KEYWORDS			
SOURCE			
ORGANISM	Porphyromonas gingivalis W83 Porphyromonas gingivalis W83 Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales; Porphyromonadaceae; Porphyromonas.		
REFERENCE	1 (bases 1 to 300454)		
AUTHORS	Nelson,K., Fleischmann,R., DeBoy,R., Paulsen,I., Fouts,D., Eisen,J., Daugherty,S., Dodson,R., Durkin,A., Gwim,M., Haft,D., Kolonay,J., Nelson,W., White,O., Mason,T., Tallon,L., Gray,J., Granger,D., Tetcelin,H., Dong,H., Galvin,J., Duncan,M., Dewhirst,F. and Fraser,C.		
TITLE	Complete Genome Sequence of the Oral Pathogenic Bacterium		
JOURNAL	Porphyromonas gingivalis Strain W83		
PUBMED	J. Bacteriol. 185 (18), 5591-5601 (2003)		
REFERENCE	12949112		
AUTHORS	2 (bases 1 to 300454)		
	Nelson,K., Fleischmann,R., DeBoy,R., Paulsen,I., Fouts,D., Eisen,J., Daugherty,S., Dodson,R., Durkin,A., Gwim,M., Haft,D., Kolonay,J., Nelson,W., White,O., Mason,T., Tallon,L., Gray,J., Granger,D., Tetcelin,H., Dong,H., Galvin,J., Duncan,M., Dewhirst,F. and Fraser,C.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-OCT-2002) The Institute for Genomic Research, 9712		
FEATURES	Medical Center Dr., Rockville, MD 20850, USA		
SOURCE	Location/Qualifiers		
	1..300454		
	/organism="Porphyromonas gingivalis W83"		
	/mol_type="genomic DNA"		
	/strain="W83"		
	/db_xref="taxon:242619"		
	177..1757		
	/locus_tag="PG0847"		
	177..1757		
	/locus_tag="PG0847"		
	/codon_start=1		
	/transl_table=11		
	/product="conserved domain protein"		
	/protein_id="AA065996.1"		
	/db_xref="GI:34396932"		
	/translation="MTDYAVKMSLDLRMTDLERSPHIDNKASHCKAKMLNDYQKQDA		
	AKGQPFVSDICTYKPGGEFNNYSEIKRLVDEYDHPSEAVFPIQCKREKAKKAVD		
	AKNRDRIITFGSTNLTGVAQQAVALIHLDLPWRSDEQRNGRAVRKGNMTAK		
	EFADKQVDVITVAVERSLDSYFENLHNQPLINDLKTTLGSRITDEGSMEDGGMN		
gene			
CDS			

[illegible]

CDS 6807..7292
/locus_tag="PG0853"
/codon_start=1
/transl_table=11
/product="DNA-binding protein, histone-like family"
/protein_id="AA066001.1"
/db_xref="GI:34396937"
/translation="MIKAIERKAGGFKTSLTWPAIHLSDVKEFEFIELVSDET
TVSSADIKAFVDRPAAYLIRLLQDGKISICGDMGTYPSTITAKAGSGVSDAVTEVL
VDKAKVYTFPRMKVTKALKGVRERARVLDVYASISIKENNNSSGSSSDNNHAG
L"
7293..8111
/locus_tag="PG0854"
7293..8111
/locus_tag="PG0854"
/note="identified by Glimmer2, putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA066002.1"
/db_xref="GI:34396938"
/translation="MYPIKIYVYKAMLDKVIDKAGGVLTAEAGVVKPKYKPIGEVL
GLPSPKICILQGMKWIINGEESIRLTSPARKIEADPAEELTEPEPVVPAQO
LGYSHSBEIRRELYANLIYSSNDVEKKGVHPFVDILKLTPTDEAKTIQYFGRDPT
EYLDLAVIKEEBGGFRITADHAKTLSDVNFEMPDNELAYLDNLVCLGIKDCGCT
FTVIDNYKLIEKRLGLSELRNQYVPEPFSIRSEKSFYQVDFGHNFTNVTVE"
complement (8333..8428)
/locus_tag="PG0855"
complement (8333..8428)
/locus_tag="PG0855"
/note="identified by Glimmer2, putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA066003.1"
/db_xref="GI:34396939"
/translation="MIFSLHFALICSEKTFVASKRAKKEVL"
8613..9518
/locus_tag="PG0856"
8613..9518
/locus_tag="PG0856"
/note="identified by Glimmer2, putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA066004.1"
/db_xref="GI:34396940"
/translation="MDENRQALASLITDITDAPIANMYLAKKVPASKNDKDLISM
IEHLNGYENLRKRYLDAVGVVDIHRGFQEVLAGYVPMGAKDKRVERLHLPIH
KSLISEVELSTKSGEGTITDIPETIWMHMRICINGDIQRAYQFATVASVKQIVAKI
KSILDIYPLKIDGESLSPFLSIKETPTQIAGIVNTGSGVNTANGATITISGANIS
INKDNPLOLORILSCIEIMKSLNSDDYKISELKEITEKDSPSNNIIRKQALQING
IAGISGVIANQVSPFLVTSAILL"
9742..9966
/locus_tag="PG0857"
9742..9966
/locus_tag="PG0857"
/note="identified by match to Pfam protein family HMM
PF01381"
/codon_start=1
/transl_table=11
/product="transcriptional regulator, putative"
/protein_id="AA066005.1"
/db_xref="GI:34396941"
/translation="MQARTIIAYVYKMKRYKLTQVDSLSEKAGVGLRFRRELQCKT
TLRLDKVKNKVFELFGSSBCGFPMKREGDDV"
9963..10298
/locus_tag="PG0858"
9963..10298
/locus_tag="PG0858"
/note="similar to GB:D90318, GB:D90319, GB:D90320,
GB:D90321, GB:D90322, GB:D90323, GB:D90324, GB:D90325,
GB:D90326, GB:D90327, GB:D90328, GB:D90329, GB:D90330,

Alignment Scores:
Pred. No.: 2, 39 Length: 300454
Score: 114.00 Matches: 26
Percent Similarity: 54.558 Conservative: 22
Best Local Similarity: 29.558 Mismatches: 40
Query Match: 22.184 Indels: 0
DB: 1 Gaps: 0
US-10-089-514-4 (1-103) x AE017175 (1-300454)
QY 3 GIUGLNASGILUENGLNARGLEUVRGALAGLUENUSPALUENUSPGLYTHIRLEUEN 22
Db 40966 CAAGACGAGAGAGCTGATCTCTTGCGCATGCAGATTGATGATGATGATGATG 41025
QY 23 ASPTHTVALARGARGARGILEASPLEUGLYVALAGLLEALARGTYRISSEARGHS 42
Db 41026 GAATTGCTAGCTCCGGCGATGCAAGTGCATACGAGATGCTTTGTTCAAAAAGACAC 41085
QY 43 GLYVALPROMETMERGINPROGLYARGVALSERLEUVALYASPARGLAALARGTYR 62
Db 41086 AATCTGGCTGTGTTAGAAATCTCCGCTACGAAACATACAGCGACCGTCCCTACT 41145
QY 63 AIAAIAASPHISGLYLEUASPGIUSERPHELEUVALSENLEUTYRASPVALIIEIETHR 82
Db 41146 GCAGCCCTCTTAGTGTGACGACAAACATTATATGCGAGCTGTTACGCCGATTCATGAG 41205
QY 83 GLUMECYARGVALGLUASPLEU 90
Db 41206 GAATCTGTCGTCGTGACACCTT 41229
RESULT 13
U67480/c 10632 bp DNA linear BCT 02-JUN-2004
LOCUS Methanocaldococcus jannaschii DSM 2661 section 22 of 150 of the
DEFINITION complete genome.
ACCESSION U67480 L77117
VERSION U67480.1 GI:2826265
KEYWORDS
SOURCE Methanocaldococcus jannaschii DSM 2661
ORGANISM Methanocaldococcus jannaschii DSM 2661
REFERENCE 1 Archaea; Euryarchaeota; Methanococci; Methanococcales;
Methanocaldococcaceae; Methanocaldococcus.
1 (bases 1 to 10632)
AUTHORS Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A., Adams,M.D.,
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.F.,
Reich,C.I., Overbeek,R., Kirtness,B.F., Weissbrock,K.G.,
Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
Fuhmann,J.L., Nguyen,D.T., Uitterlbeck,T., Kelley,J.M.,
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Frazer,C.M.,
Smith,H.O., Woese,C.R. and Venter,J.C.
TITLE Complete genome sequence of the methanogenic archaeon,
Methanococcus jannaschii
JOURNAL Science 273 (5278), 1058-1073 (1996)
MEDLINE 96337999
PUBMED 8686087
REFERENCE 2 (bases 1 to 10632)
AUTHORS Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
Sutton,G.G., Blake,J.A., Fitzgerald,L.M., Clayton,R.A.,
Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.F.,
Adams,M.D., Reich,C.I., Overbeek,R., Kirtness,B.F., Weissbrock,K.G.,
Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
Fuhmann,J.L., Nguyen,D.T., Uitterlbeck,T., Kelley,J.M.,
Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,

Roberts, K.M., Kaine, B.B., Borodovsky, M., Klenk, H.P., Fraser, C.M.,
Smith, H.O., Moese, C.R. and Venter, J.C.
Direct Submission
Submitted (27-AUG-1996) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jan 30, 1998 this sequence version replaced gi:1590974.
Location/Qualifiers
FEATURES
source

1. 10632
/organism="Methanocaldococcus jannaschii DSM 2661"
/mol_type="genomic DNA"
/strain="DSM 2661"
/db_xref="taxon:243232"
complement (59. .127)
/locus_tag="MJ0242"
complement (59. .127)
/locus_tag="MJ0242"
/note="similar to SP:P28751 percent identity: 63.64;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="LSU ribosomal protein L41E"
/protein_id="AAB98230.1"
/db_xref="GI:1592277"
/translation="MRKWKYKRLRLRKRKRARS"
complement (244. .492)
/locus_tag="MJ0243"
complement (244. .492)
/locus_tag="MJ0243"
/note="similar to GB:A500666 percent identity: 47.95;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB98231.1"
/db_xref="GI:1590976"
/translation="MDKTYEKIKKEAEIINKFSEVLEKFNLEMESEYYIIDTRNVL
RDEAVSNPEPEKFLKIAPKVKEGVVVEKSWLK"
complement (511. .1380)
/locus_tag="MJ0244"
complement (511. .1380)
/locus_tag="MJ0244"
/note="similar to GB:L08471 SP:Q04796 PID:142830
GB:A1009126 percent identity: 45.64; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="dihydrodipicolinate synthase (dapA)"
/protein_id="AAB98232.1"
/db_xref="GI:1590977"
/translation="MEKGVPAITTPKKNENDPDGLEENINPLIENGSGIYAVGT
GESPTLSHEERKVIKQVDVNGRVVINGAGSNCETBEAIEISVFAEDVGADAVLSI
TPYNNKPTQGLRKHFGKVAESINLPVLVNGPSRTAVNLEPKVLAEYSNLSAV
KEANPNLSQVSELIHDAKITVLGNDLTPILALGGKGVSVVAVNPVEFEMVY
ALGDPEKAREIHVYKLEPLMKAMFIETNPVPKVALNMMGRPAGELRLPLCEMSEHK
KILENVKDLGLT"
complement (1391. .1582)
/locus_tag="MJ0245"
complement (1391. .1582)
/locus_tag="MJ0245"
/note="similar to SP:P02407 SP:P14127 PID:172458
PID:172460 PID:575695 percent identity: 51.61; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="SSU ribosomal protein S17E"
/protein_id="AAB98233.1"
/db_xref="GI:1590978"
/translation="MMGRIRQTLIKRTAMELIKRYRDLFTTDFETNKRVLEVAQIST
KRLRNRTAGIITHMRQLQ"
complement (1627. .1926)
/locus_tag="MJ0246"
complement (1627. .1926)
/locus_tag="MJ0246"

/note="similar to SP:P07022 PID:147175 PID:551824
GB:U00096 PID:1788951 percent identity: 34.78; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chorismate mutase/prephenate dehydratase (pheA)"
/protein_id="AAB98234.1"
/db_xref="GI:1590979"
/translation="MTEKAEIRKTIKIDINKTIKLAEKNSIAKVAETKQGLPI
NDPERKTIYDRKLCKEHVDENIGIKIPLLIRHAKALQVLEETONKAKK"
2228. .2971
/locus_tag="MJ0247"
2228. .2971
/locus_tag="MJ0247"
/note="similar to SP:P15873 PID:4193 PID:476047 PID:536356
percent identity: 31.33; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="proliferating-cell nuclear antigen (p130)"
/protein_id="AAB98235.1"
/db_xref="GI:1590980"
/translation="MFRGVMSAKERKVVDTISTLLDEICEVEDEGIKASAMDSH
VALVSLIEIRLAFEEYADSHDIGIDLEAFKKVMNKAADRLILDESKNRLNVI
EMTKRKFSLALDIISSASVKEIYEPVAVIMKGAFAKALQADLFSFYVILKADE
DKRVTHAKGDINENEAIPKDSASISLEVKBAKSAFNLDIYIMDVKVGSSGDIKI
YQNDMPLEKYSIAGVNLFLAPRIEG"
2999. .3583
/locus_tag="MJ0248"
2999. .3583
/locus_tag="MJ0248"
/note="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="M. jannaschii predicted coding region MJ0248"
/protein_id="AAB98243.1"
/db_xref="GI:1590981"
/translation="MRVPTIIYFRFCSKDHKRYPIYTMYSISKYFPFEIEKNDKLLK
LPDPFYDDIRVYKIKDIDLELRVRYFKELKLTLYALVYDNERNLPELNIIT
HAIENIVELKIEETPEPKPTEIDTPKPIYINDVVKDNPFPDPGTPIYDLN
KNDVLSIDRKISIIIEKHRIISRIGSYENPERS"
3561. .3845
/locus_tag="MJ0249"
3561. .3845
/locus_tag="MJ0249"
/note="similar to SP:P32411 percent identity: 38.71;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="LSU ribosomal protein L44E"
/protein_id="AAB98236.1"
/db_xref="GI:1499028"
/translation="MKIPKVRVRCYCKKTHIIVEKAKKPSLELTWGORFRRT
AGYGFPRPLDPSKRVKIDLRPKCTEGKMTKANGCFRSGRPFVEK"
3913. .4101
/locus_tag="MJ0250"
3913. .4101
/locus_tag="MJ0250"
/note="similar to SP:P35997 SP:P38711 PID:213116
PID:486271 PID:500694 percent identity: 42.59; identified
by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="SSU ribosomal protein S27E"
/protein_id="AAB98237.1"
/db_xref="GI:1499029"
/translation="MMELIPQRTKRLRVQCEBNNBOIVFGSPATVAKCLTCGKVLV
EPRGGKGVAKIKLELTG"
complement (6214. .6657)
/locus_tag="MJ0251"
complement (6214. .6657)
/locus_tag="MJ0251"

/note="similar to SP:P00211 percent identity: 37.04;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="ferredoxin II"
/protein_id="AB98238.1"
/db_xref="GI:1499030"
/translation="MENLEKIEELKKIREFLINLEIKKMOELNVDSDIYAEYKV
TKIVREPNIKLYQYDAIKEMFEYEEKRKDIISWPKIDVRCNKCKISFCPRG
VYDAENGKVVVKKPKYSCIVNCNCSIMCCENNAIIFPDKIPRRN"
6715..7419
/locus_tag="MJ0252"
6715..7419
/locus_tag="MJ0252"
/note="similar to SP:P42405 PID:710637 PID:1805418
GB:AL009126 percent identity: 34.48; identified by
sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AB98239.1"

Alignment Scores:

Pred. No.: 0.114 Length: 10632
Score: 112.00 Matches: 20
Percent Similarity: 53.09% Conservative: 23
Best Local Similarity: 24.69% Mismatches: 38
Query Match: 21.79% Indels: 0
DB: 1 Gaps: 0

US-10-089-514-4 (1-103) x U67480 (1-10632)

QY 1 MetThrGluGlnAspGluLeuArgValArgValSerLeuValAspGlyThr 20
Db 1932 GTTACATGATAGAGAACTCTCGAATTAAGAGAAAGTTATGATGCAATTAAG 1873
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSer 40
Db 1872 ATATTAAAGCTAATTGCTGAAGAAATATGTTAGCTAAGAGTACGTGATTAATAAT 1813
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValAspArgAlaAla 60
Db 1812 CAGCTTGATTCCTATTAAAGCAAGCAAGAAATATATATATACGATGATTAAGTAAG 1753
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
Db 1752 AAACCTTGTAAGACATTAACGTTGATGAATAATTTGGCATTTAAATATTCAATTAATT 1693
QY 81 Ile 81
Db 1692 ATA 1690

RESULT 14
AR271569_12
WPCOMMENT

Sequence split into 17 fragments LOCUS AR271569 Accession AR271569

Fragment Name	Begin	End
AR271569_00	1	110000
AR271569_01	100001	210000
AR271569_02	200001	310000
AR271569_03	300001	410000
AR271569_04	400001	510000
AR271569_05	500001	610000
AR271569_06	600001	710000
AR271569_07	700001	810000
AR271569_08	800001	910000
AR271569_09	900001	1010000
AR271569_10	1000001	1110000
AR271569_11	1100001	1210000
AR271569_12	1200001	1310000
AR271569_13	1300001	1410000
AR271569_14	1400001	1510000
AR271569_15	1500001	1610000

AR271569_16 1600001 1664976
Continuation (13 of 17) of AR271569 from base 1200001 (AR271569 Sequence 1 from patent L

Alignment Scores:

Pred. No.: 1.36 Length: 110000
Score: 112.00 Matches: 20
Percent Similarity: 53.09% Conservative: 23
Best Local Similarity: 24.69% Mismatches: 38
Query Match: 21.79% Indels: 0
DB: 6 Gaps: 0

US-10-089-514-4 (1-103) x AR271569_12 (1-110000)

QY 1 MetThrGluGlnAspGluLeuArgValArgValSerLeuValAspGlyThr 20
Db 38358 GTTACATGATAGAGAACTCTCGAATTAAGAGAAAGTTATGATGCAATTAAG 38417
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSer 40
Db 38418 ATATTAAAGCTAATTGCTGAAGAAATATGTTAGCTAAGAGTACGTGATTAATAAT 38477
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValAspArgAlaAla 60
Db 38478 CAGCTTGATTCCTATTAAAGCAAGCAAGAAATATATATATACGATGATTAAGTAAG 38537
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
Db 38538 AAACCTTGTAAGACATTAACGTTGATGAATAATTTGGCATTTAAATATTCAATTAATT 38597
QY 81 Ile 81
Db 38598 ATA 38600

RESULT 15
AE000858/c

LOCUS AE000858 13242 bp DNA linear BCT 19-JUN-2002
DEFINITION Methanobacterium thermoautotrophicum from bases 725908 to 739149
(section 64 of 148) of the complete genome.
ACCESSION AE000858 AE000666
VERSION AE000858.1 GI:2621885

KEYWORDS
SOURCE

ORGANISM

REFERENCE

AUTHORS

Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H., Dubois,J.,
Aldredge,T., Bashirzadeh,R., Blakey,D., Cook,R., Gilbert,K.,
Harrison,D., Hoang,L., Keagle,P., Lum,W., Pothier,B., Qiu,D.,
Spadafora,R., Vicaire,R., Wang,Y., Wierzbowski,J., Gibson,R.,
Jiwani,N., Caruso,A., Bush,D. and Reeve,J.N.
Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics
J. Bacteriol. 179 (22), 7135-7155 (1997)

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..13242
/organism="Methanobacter thermotrophicus str.
Delta H"
/mol_type="genomic DNA"
/strain="Delta H"
/db_xref="taxon:187420"
/clone="WTH"
/note="synonym: Methanobacterium thermoautotrophicum str.
deltaH"

gene	62. .676	/gene="MTH796"	/product="dihydrodipicolinate reductase"
CDS	62. .676	/protein_id="AAB85300.1"	/db_xref="GI:2621890"
	/gene="MTH796"	/note="Function Code:14.01 - Unknown, Conserved protein; similar to, pIR:LN:G64397 AC:G64397, p()=5.5E-17, pid=30%"	
	/codon_start=1	/transl_table=11	
	/product="conserved protein"		
	/protein_id="AAB85296.1"		
	/db_xref="GI:2621886"		
	/translation="MKIIDLTHKIEDSPVPFGDDPPVYLARI ESDADYITSSLSLGNH TGRHIDPLAHGCSALTVDGIGLEELTGEGLVSGGSIINGGDIAVIRGTSMSSWGSE GYFRDYFGIKRLAEIVHEHVGCVDPSPVDRPGEDTHKLLKXGIVIVENTITVT DILPKPRLFVPLPSVAEASPARVFAVTSGSGSSGSSSDRSSFQR"		
gene	complement (680. .1216)		
CDS	/gene="MTH797"		
	/complement (680. .1216)		
	/note="Function Code:14.00 - Unknown, ; similar to, pIR:LN:S33068 AC:S33068, p()=0.055, pid=34%"		
	/codon_start=1		
	/transl_table=11		
	/product="unknown"		
	/protein_id="AAB85297.1"		
	/db_xref="GI:2621887"		
	/translation="MERDTMRFRDRDSRKEQKULYKACMEIETKLRDLAKCRNSDN CERVELDEDLKTRDRIDELERIMAKEDRVIOELKGIADKESRITDLERYEYFQAL YTGKPEKDLTSFQSIYRLLPDERATTEMLDFINIGFDLKLNNVOLLRLERKGL YFRSVSEGRRTLMKVKR"		
gene	complement (1431. .1991)		
CDS	/gene="MTH798"		
	/complement (1431. .1991)		
	/note="Function Code:12.09 - Cell Processes, Protein and peptide secretion ; similar to, sp:LN:Y374_METUA AC:Q57819, p()=1.7E-13, pid=28%"		
	/codon_start=1		
	/transl_table=11		
	/product="bacitracin transport permease related protein"		
	/protein_id="AAB85298.1"		
	/db_xref="GI:2621888"		
	/translation="MHPTLLGPDLSVLVYFNFSVLAAPFLDPLMLITFGGTAFWVL CILVILGSEDEREAAFMALTAIVGSELSALKVYARPRYEVIGMVAHNAVAGY SMESGAAVAAPAGISLYFRLGRPWFTIILASLVGSRITVGLHYPSDVLAGAVLGLV CAPTALKIDERVCFGLCRFDRLQON"		
gene	complement (1998. .3041)		
CDS	/gene="MTH799"		
	/complement (1998. .3041)		
	/note="Function Code:5.01 - L-Amino Acid Metabolism, Alanine--aspartate and glutamate metabolism ; similar to, sp:LN:DHA5_METUA AC:Q57658, p()=2.4E-96, pid=55%"		
	/codon_start=1		
	/transl_table=11		
	/product="aspartate-semialdehyde dehydrogenase"		
	/protein_id="AAB85299.1"		
	/db_xref="GI:2621889"		
	/translation="MVNVGLGATGVGQRFIOMLDKHPFELITLAASRSAGKPYG EVANWYLDCEPESVDMEVETDPSAVGVDILFSLPADVARKVPEKPAKCYIAS NASAMMEPPLVPIEVPNEPLDLEVOGRRGSGFITVNPNCSTIALITLTKPY DAYTIKRVVSTQAVSGAGYNGVPSMAIINDNVPRFSGSEKILETFTILLSELDBG VKPAPFVGVSASCHRVFVVDGHTBAVFTELDDEFDIDVREANDKFRGLQKGLHSA PEKPVAVRDEENRPQPMRDMDGMAVTVAGRLREDAAPFNSILRYLVGHNTVGAAG ASILNLELNEIL"		
gene	complement (3054. .3875)		
CDS	/gene="MTH800"		
	/complement (3054. .3875)		
	/note="Function Code:5.07 - L-Amino Acid Metabolism, lysine biosynthesis ; similar to, pIR:LN:F64352 AC:F64352, p()=8E-80, pid=56%"		
	/codon_start=1		
	/transl_table=11		
	/product="aspartate kinase II alpha subunit"		
	/protein_id="AAB85302.1"		
	/db_xref="GI:2621892"		
	/translation="MELIVAKFGGTSINGRRIKKARSVYKMKRKYVVVSAIN KTTDELQIVDEAMEDVATEKQIAEIVSGMESVSI FFSALVIALCVKSYIDPFME WPIITDSNLNAKVDPEATEBEKSRBELLLDGOIIPVCGFLGRDNGYTTTGSGS DITAPFLGICAKADDEVITVDGSGVSTPNTKQKAKDKLSVEEMRDLATGAGVLL HPAALTKXPDPDKAKIIGFBHGLSAPGFIPTPSKRYKKTITLDPDPSVAVVGS KINKKGIARLTSRLAENSINITIGISTGNSVTITVDKXDAEARLRLHVDVIADD LSLSLGRDIAMTITSSPFDITPGIISBTYKLRDNDINIVISSQTSVIVFVDMN DGKAYELVRGVLE"		
gene	complement (6046. .6234)		
CDS	/gene="MTH803"		
	/complement (6046. .6234)		
	/note="Function Code:10.04 - Metabolism of Macromolecules, Ribosomal proteins; similar to, sp:LN:R17E_METUA AC:P54026, p()=3.3E-12, pid=46%"		
	/codon_start=1		
	/transl_table=11		
	/product="ribosomal protein S17"		
	/protein_id="AAB85303.1"		
	/db_xref="GI:2621893"		
	/translation="MGNIRTSFVKRIAKEMIETHPEKFTDDPDTNKKLVBEFSTGSK HLRNKAGIYITISQK"		
gene	complement (6236. .6532)		
CDS	/gene="MTH804"		
	/complement (6236. .6532)		
	/note="Function Code:5.11 - Tyrosine metabolism ; similar to, pIR:LN:G64330 AC:G64330, p()=8.9E-10, pid=32%"		
	/codon_start=1		
	/transl_table=11		
	/product="chorismate mutase, subunit A"		
	/protein_id="AAB85304.1"		
	/db_xref="GI:2621894"		
	/translation="MDEVARAEVLRSRQKIDGRDIDILITSRIALAREIAKEV		

gene
complement (6573 . .7442)
/gene="MT1805"
complement (6573 . .7442)
/gene="MT1805"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, p1:LN:G64479 AC:G64479, p1=1.8E-36, pid=31%"
/codon_start=1
/transl_table=11
/product="conserved protein (homoserine kinase related)"
/protein_id="AAB85305.1"
/db_xref="GI:2621895"
/translation="MNVVKTVRSPGSATVYNAIATGSGAFGIGLRVEAEALIDSG
VECSIRSGADTSLMELCTRMVIEHVGVDGVRVTSDDLPAVSGLSSSAASNAVTMA
VSSLSDEFGIEMEDFEMLMNMAVDASLQAGVSVTGAYDDASAFYGLITVDNMERR
ILIREPMENQKVLIVMPDRKSLTAQSDVPRMKLLAPWDMAPREVLDGRVHSALTNG

Alignment Scores:
Pred. No.: 0.268 Length: 13242
Score: 109.50 Matches: 26
Percent Similarity: 54.55% Conservative: 22
Best Local Similarity: 29.55% Mismatches: 39
Query Match: 21.30% Indels: 1
DB: 1 Gaps: 1

US-10-089-514-4 (1-103) x AE000858 (1-13242)

QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
Db 6505 CTCAGAGATCCGACGACGATAGATGATGACGACGACGACATCTGACCTCATTAACC 6446
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSerArgHisGlyValProMet 46
Db 6445 TCAAGGATACCCCTTGACGCGGATAGACGAGGCCAAGAGGTCCTTGGATGAGATC 6386
QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66
Db 6385 CTGGATCCTGAAAGGAACTCCAGATTATAGAGAGACCCGAAATTCGCCAGGAAAT 6326
QY 67 GlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 86
Db 6325 GGATCGATGAGATAAATTACAGAACTGATGAGATTCTGATG--GATTCAGTAA 6269
QY 87 ValGluAspLeuValMetSerArg 94
Db 6268 ACTGAACAGAAAGAAATGTTAAG 6245

Search completed: October 5, 2005, 15:30:35
Job time: 1687.4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 4, 2005, 21:38:23 ; Search time 180.876 Seconds
(without alignments)
3371.006 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514
Sequence: 1 MFEQNELQRLAEALDGT.....MCRVEDLVMSRESLTAEDDR 103

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+pn.model -DEV=rlh
-Q=/cgn2_1/USPTO.spool_h/US10089514/runat_04102005_105743_8027/app.query.fasta_1.1621
-DB=N_Geneseq -QMT=fastlap -SUFFIX=ring -MINMATCH=0.1 -LOOFC=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blcsm62 -TRANS=human40.cdl -LIST=45
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pro -NORR=ext -HEADSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10089514 @CCN_1.1.860 @runat_04102005_105743_8027 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

N_Geneseq_16Dec04:*
1: geneseqn1960s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	312	4	AAf86410
2	514	100.0	312	8	AAf86410 4-amino-4
3	486.5	94.6	3305	8	AAf86410 1 S venezue
4	439	85.4	12391	8	ABZ69799 Plasmid P
5	256	49.8	387	2	AAf86410 1 S venezue

6	256	49.8	2888	2	AAT59268	AAT59268 Streptococ
7	120	23.3	360	10	ACF71772	ACF71772 Phototrab
8	120	23.3	110000	10	ACF67367_49	Continuation (50 o
9	120	23.3	110000	10	ACF65387_1	Continuation (2 of
10	112	21.8	110000	2	AAW21209_12	Continuation (13 o
11	109	21.2	94968	11	ADM27081_16	Continuation (17 o
12	106	20.6	1125	10	ADF00853	Adf00853 Bacteri
13	105	20.4	1305	11	ACF98708	ACF98708 Klebsiell
14	102	19.8	1122	8	AAf86410	AAf86410
15	102	19.8	1122	12	ADJ98178	ADJ98178 Pantoea
16	102	19.8	1122	6	ABN83449	ABN83449 Erwinia
17	101	19.6	1253	1	AAW71054	AAW71054 Optimized
18	101	19.6	2058	1	AAW71053	AAW71053 Escherich
19	101	19.6	3659	1	AAW71109	AAW71109 phea arof
20	101	19.6	4189	1	AAW71111	AAW71111 phea arof
21	96.5	18.8	1833	6	AAf86410	AAf86410
22	96	18.7	273	6	ABN66826	ABN66826 Streptoc
23	96	18.7	110000	6	ABN71527_05	Continuation (6 of
24	94	18.3	339	6	ABN69596	ABN69596 Streptoc
25	93	18.1	1131	11	ACH98777	ACH98777 Klebsiell
26	92	17.9	1122	10	ACF69143	ACF69143 Phototrab
27	92	17.9	110000	10	ACF67367_20	Continuation (21 o
28	92	17.9	243072	10	ACF65382	ACF65382 Phototrab
29	90	17.5	649	10	ACF66455	ACF66455 Phototrab
30	90	17.5	1122	6	ABN83448	ABN83448 Escherich
31	90	17.5	1122	8	AAf86410	AAf86410
32	90	17.5	1158	10	ACF69141	ACF69141 Phototrab
33	90	17.5	1238	6	ABN99612	ABN99612 E. coli
34	90	17.5	1238	6	AAf86410	AAf86410
35	90	17.5	110000	10	ACF67367_20	Continuation (21 o
36	90	17.5	243072	10	ACF65382	ACF65382 Phototrab
37	89.5	17.4	1104	8	ACA27476	ACA27476 Prokaryot
38	89.5	17.4	18096	4	AAf86410	AAf86410
39	89.5	17.4	18096	8	ACF64517	ACF64517 Phototrab
40	89	17.3	652	3	AAf86410	AAf86410
41	89	17.3	652	6	ABN18494	ABN18494 Human ORF
42	88.5	17.2	804	11	ABD06101	ABD06101 Pseudomon
43	88.5	17.2	960	11	ABD06101	ABD06101 Pseudomon
44	88.5	17.2	1653	11	ABD06120	ABD06120 Pseudomon
45	88.5	17.2	1791	11	ABD06080	ABD06080 Pseudomon

ALIGNMENTS

RESULT 1	AAf86410	standard; DNA, 312 BP.
ID	AAf86410	
XX	AAf86410	
AC	AAf86410	
XX	AAf86410	
DT	22-JUN-2001	(first entry)
XX	22-JUN-2001	
DE	4-amino-4-deoxychorismic acid mutase coding sequence.	
XX	4-amino-4-deoxychorismic acid mutase coding sequence.	
KM	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;	
KW	4-amino-4-deoxychorismic acid mutase; enzyme; pabB; ds.	
XX	4-amino-4-deoxychorismic acid mutase; enzyme; pabB; ds.	
OS	Streptomyces venezuelae.	
XX	Streptomyces venezuelae.	
FT	Key	Location/Qualifiers
FT	CDS	1..312
FT		/tag= a
FT		/product= "4-amino-4-deoxychorismic acid mutase"
XX	WO200123542-A1.	
XX	05-APR-2001.	
XX	29-SEP-2000; 2000WO-DP06783.	
XX	29-SEP-1999; 99JP-00276314.	
XX	(MEIJ) MEIJI SEIKA KAISHA LTD.	

XX Yanai K, Okakura K, Yaeuda S, Watanabe M, Miyamoto K, Midoh N;
PI Murakami T;
XX
DR WPI: 2001-290517/30.
XX P-PSDB; AAB82072.
XX
PT Transformant producing secondary metabolite modified with functional
PT group e.g. benzene with nitrogen-containing substituent at para-position,
PT PF1022, with ease at low cost, for application in pharmaceuticals and
PT agrochemicals.
XX
PS Claim 13; Page 65; 83pp; Japanese.
XX
XX The present invention relates to a transformant having been modified so
CC as to produce a secondary metabolite. The secondary metabolite has a
CC benzene ring skeleton free from substitution at the para-position by a
CC nitrogen-containing functional group, thereby enabling the production of
CC a secondary metabolite with a benzene ring skeleton substituted at the
CC para-position by a nitrogen-containing group. The transformant organism
CC of the present invention has been produced by transferring a gene
CC participating in the biosynthesis pathway from chorismic acid into p-
CC aminophenylpyruvic acid. The present sequence is the coding sequence for
CC 4-amino-4-deoxychorismic acid mutase (pabB), from Streptomyces
CC venezuelae. pabB participates in the biosynthesis pathway from chorismic
CC acid into p-aminophenylpyruvic acid, and so the pabB gene can be used to
CC produce the transformant of the present invention. The transformant can
CC be used to produce metabolites for application in pharmaceuticals,
CC veterinary drugs and agrochemicals
XX
SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,38e-61 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-089-514-4 (1-103) x AAF86410 (1-312)
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 1 ATGACCGAGCAGACAGAGCTGACGCGGCTGCGCGGAGCTCGACCGCTCGACGGGACG 60
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSer 40
DB 61 CTCCTGGACACGGTGGCGGCGCGCATCGACTCGGTCCGCGATCGCGGGTACAGTCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAlaAla 60
DB 121 CGGCAAGGGGTCCGATGATGACGCCCGCGGCTAGCTGTCAAGACAGGCGCGCC 180
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
DB 181 CGCTACGGCGCGCAGCACCGGCTCGACGAACTCTCGGTGAACCTTACGACGTGATC 240
QY 81 IleThrGluMetCysArgArgValGluAspLeuValMetSerArgGluSerLeuThrAlaGlu 100
DB 241 ATTCAGGAATGTGGCGCGTGGAGACCTGTGTATGACCGGAGAGCCTGACGGCGCAG 300
QY 101 AspArgArg 103
DB 301 GACCGGCGG 309
RESULT 2
AAL50181
ID AAL50181 standard; DNA; 312 BP.
XX
AC AAL50181;
XX
DT 28-JAN-2003 (first entry)

XX
DE S venezuelae PF1022 substance gene #2.
XX
XX Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;
KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
KW veterinary drug; gene; ds.
XX
XX Streptomyces venezuelae.
XX
FH Key Location/Qualifiers
FT CDS 1..312
FT /tag=a
FT /product="PF1022 substance"
XX
XX WO200277244-A1.
XX
XX 03-OCT-2002.
XX
XX 22-MAR-2002; 2002WO-JP002782.
XX
XX 22-MAR-2001; 2001JP-00082227.
XX
XX (MEIJU) MEIJU SEIKA KAISHA LTD.
XX
XX Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;
PI
DR WPI: 2003-018934/01.
DR P-PSDB; AAO19564.
XX
XX Novel biosynthesis gene-transferred transformants for producing PF1022
PT substance derivatives by fermentation, as pharmaceuticals or veterinary
PT drugs with anthelmintic activity.
XX
XX Claim 15; Page 85; 116pp; Japanese.
XX
XX The present invention relates to transformants capable of producing
CC PF1022 substance derivatives. These were obtained by transferring a gene
CC participating in the biosynthesis pathway from chorismic acid to p-
CC aminophenylpyruvic acid (biosynthesis gene) into a phenylalanine-
CC requiring host derived from an organism producing the PF1022 substance.
CC The transformants are producing PF1022 substance derivatives by
CC fermentation, for use as pharmaceuticals or veterinary drugs. The present
CC sequence is a substance PF1022 coding sequence from Streptomyces
XX venezuelae
SQ Sequence 312 BP; 54 A; 106 C; 111 G; 41 T; 0 U; 0 Other;
XX
Alignment Scores:
Pred. No.: 2,38e-61 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: Gaps: 0
US-10-089-514-4 (1-103) x AAL50181 (1-312)
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 1 ATGACCGAGCAGACAGAGCTGACGCGGCTGCGCGGAGCTCGACCGCTCGACGGGACG 60
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSer 40
DB 61 CTCCTGGACACGGTGGCGGCGCGCATCGACTCGGTCCGCGATCGCGGGTACAGTCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAlaAla 60
DB 121 CGGCAAGGGGTCCGATGATGACGCCCGCGGCTAGCTGTCAAGACAGGCGCGCC 180
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
DB 181 CGCTACGGCGCGCAGCACCGGCTCGACGAACTCTCGGTGAACCTTACGACGTGATC 240

Qy 81 |||lethrgluMetCysaRgVaIgluaAspleuValMetSerArgIuSerIeuThrAlaGlu 100
 |||
 Db 241 ATACCGAGATGTCCTCCGCTGAGAGACTGTGATGACCCGAGAGCCTTACGCGCCAG 300
 Qy 101 |||AspArgArg 103
 |||
 Db 301 GACCGCGCG 309

RESULT 3
 ABZ69799
 ID ABZ69799 standard; DNA; 3305 BP.
 XX
 AC ABZ69799;
 XX
 DT 08-APR-2003 (first entry)
 XX
 DE Plasmid papABC.
 XX
 KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
 KM biosynthesis; p-aminophenylalanine; paf; choriassmate.
 XX
 OS Synthetic.
 OS
 PN W0200285923-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002WO-US012465.
 XX
 PR 19-APR-2001; 2001US-0285030P.
 PR 06-FEB-2002; 2002US-0355514P.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
 PI Meggers EL, Wehl RA, Pasternak M, Santoro SW, Zhang Z;
 DR WPI; 2003-120430/11.
 XX
 PT Composition useful for producing protein comprising unnatural amino acid,
 PT has translation system comprising orthogonal tRNA and orthogonal
 PT aminoacyl tRNA synthetase.
 PS
 PS Example 4; Page 127-128; 188bp; English.
 XX
 CC The invention relates to a novel composition comprising a translation
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
 CC -tRNA with at least one unnatural amino acid in the translation system
 CC and the O-tRNA recognises at least one selector codon. A composition of
 CC the invention is useful for producing at least one protein comprising at
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TNG mutant of mouse
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
 CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
 CC acid is provided exogenously. The translation system is a cell and the
 CC unnatural amino acid is biosynthesised by the cell. The present sequence
 CC represents a plasmid containing the individual genes papABC that encode
 CC the enzymes used to carry out the conversion of choriassmate to the
 CC unnatural amino acid p-aminophenylalanine (paf)
 XX
 SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,696-56 Length: 3305
 Score: 486.50 Matches: 103
 Percent Similarity: 98.10% Conservative: 0
 Best Local Similarity: 98.10% Mismatches: 2
 Query Match: 94.65% Indels: 0
 DB: 8 Gaps: 1

US-10-089-514-4 (1-103) x ABZ69799 (1-3305)
 Qy 1 MetThrIuGluInuEngIuLeuGln---ArgIeuAraGAlaGluIeuAspAlaIeuAspG 19
 |||
 Db 2990 ATGACCGAGCAGACGAGAGTGCAGGCTTGCCTGCGCGGAGCTCGACCCCTCGACGG 3049
 Qy 19 |||ThrIeuIeuAspThrValAraArgArgIleAraIeuGlyValArgIleAlaArgTyrI 39
 |||
 Db 3050 GACGCTTCTGACACGGGTGGCGCGCGCATCGACCTCGGTGTCGACATCGCGGATCAAA 3109
 Qy 39 sSerArgIleGlyValPrometMetGlnProGlyArgValSerIeuValIysAspArgAl 59
 |||
 Db 3110 GTCCCGGACGAGGCTCCCGATGATGACGCCCGCGGCTCAGCTGTTCAAGACACAGGC 3169
 Qy 59 aAlaArgTyrAlaAlaAspHisGlyIeuAspGluSerPheIeuValAsnIeuTyrAspVa 79
 |||
 Db 3170 CGCCCGCTACCGCGCGCACCGCTCGACGATCTTCTGTTGAACTCTACGACGCT 3229
 Qy 79 |||lethrgluMetCysaRgVaIgluaAspleuValMetSerArgIuSerIeuThrAl 99
 |||
 Db 3230 GATCATCAGGAGATGTCCTCCGCTGAGAGCTGTGTGATGAGCCGGAGAGCCTGACGCG 3289
 Qy 99 aGluAspArgArg 103
 |||
 Db 3290 CGAGGACCGCGCG 3302

RESULT 4
 ABZ69798
 ID ABZ69798 standard; DNA; 12391 BP.
 XX
 AC ABZ69798;
 XX
 DT 08-APR-2003 (first entry)
 XX
 DE Plasmid plasc-papabc.
 XX
 KM DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
 KM orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
 KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
 KM biosynthesis; p-aminophenylalanine; paf.
 XX
 OS Synthetic.
 OS
 PN W0200285923-A2.
 XX
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002WO-US012465.
 XX
 PR 19-APR-2001; 2001US-0285030P.
 PR 06-FEB-2002; 2002US-0355514P.
 XX
 PA (SCRI) SCRIPPS RES INST.
 XX
 PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;
 PI Meggers EL, Wehl RA, Pasternak M, Santoro SW, Zhang Z;
 DR WPI; 2003-120430/11.
 XX
 PT Composition useful for producing protein comprising unnatural amino acid,
 PT has translation system comprising orthogonal tRNA and orthogonal
 PT aminoacyl tRNA synthetase.
 PS
 PS Example 4; Page 124-127; 188bp; English.
 XX
 CC The invention relates to a novel composition comprising a translation
 CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
 CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
 CC -tRNA with at least one unnatural amino acid in the translation system
 CC and the O-tRNA recognises at least one selector codon. A composition of
 CC the invention is useful for producing at least one protein comprising at
 CC least one unnatural amino acid. The protein is the Asp127AG mutant of
 CC chloramphenicol acetyltransferase (CAT), the Tyr163TNG mutant of mouse

CC dihydrofolate reductase (DHFR), or a Tyr163TNG mutant of mouse
CC dihydrofolate reductase comprising CHOOH His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine
CC (pAp) in vivo

XX SQ Sequence 12391 BP, 2830 A, 3588 C, 3346 G, 2627 T, 0 U, 0 Other;

Alignment Scores:

Pred. No.:	7,79e-49	Length:	12391
Score:	439.00	Matches:	96
Percent Similarity:	94.12%	Conservative:	0
Best Local Similarity:	94.12%	Mismatches:	3
Query Match:	85.41%	Indels:	3
DB:	8	Gaps:	2

US-10-089-514-4 (1-103) x ABZ69798 (1-12391)

OY 1 MetThrGluGlnAsnGluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG1 19
DB 2055 ATGACCGGACAGAACGAGCTGCAGGTTCGGCGCTGCCGCGAGCTCGACCCCTCGACGG 2114
OY 19 YhrLeuLeuAspThrValArgArgArgTleAspLeuGlyValArgTleAlaArgTyr1 39
DB 2115 GACGCTTCGGACACGCGTGGCGCGCATCGACCTCGGTTCGCGATCGCGGATACAA 2174
OY 39 sSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLysAspArg1 59
DB 2175 GTCCCGGACGCGGTCCCATGATGACACCCGCGGCTGACCTGTCAAGACAGGGC 2234
OY 59 aAlaArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspVa 79
DB 2235 CGCCCGCTACGCGCGGACGACCGCTCGCATGCTTCGTGTGAACCTCTACGACGT 2294
OY 79 lIleIleThrGluMetCysArgValAluAspLeuValMetSerArgGluSerLeuThr1 99
DB 2295 GATCATCAGGAGATGTGCCGCGTGAAGACCTGTGATGAGC---CCGTCACTACTAA 2351
OY 99 aGlu 100
DB 2352 GGAG 2355

RESULT 5
AAT58552 standard; cDNA; 387 BP.

XX AC AAT58552;

DT 02-APR-1997 (first entry)

DE Streptomyces pristinaespiralis papB gene.

XX Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
KM DMPAPa precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
KW papC; isomerisation; aromatization; N-methyltransferase; ds.

OS Streptomyces pristinaespiralis.

XX Key Location/Qualifiers
FH mat_peptide 1..387
FT /*tag= a
FT /product= "PapB"

PN MO9601901-A1.

PD 25-JAN-1996.

PF 04-JUL-1995; 95MO-FR000889.

PR 08-JUL-1994; 94FR-00008478.

XX (RHON) RHONE POUJENC RORER SA.

XX Blanc V, Thibaut D, Bamae-Jacques N, Blanche F, Crouzet J;
PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
XX WPI; 1996-097631/10.

DR P-PSDB; AAM11583.

PT New streptogramin B derivs. useful as antibiotics - produced by new
PT mutants of Streptomyces having altered genes for streptogramin B
PT biosynthesis.

PS Claim 18; Page 106-107; 146pp; French.

XX The papA and papM genes of S.pristinaespiralis are involved in the
CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPa), a precursor for
CC pristinamycin IA. The region between these two genes was sequenced and
CC two open reading frames were identified. The first (papC) was on the
CC complementary strand and decodes to an amino acid sequence with homology
CC to the region of E.coli TyrA which has been implicated in aromatization
CC reactions. The papC gene product is likely to be involved in a similar
CC aromatization of 4-deoxy 4-amino prephenate to give 4-amino
CC phenylpyruvate during DMPAPa synthesis. The second open reading frame
CC (papB) could be decoded to give a product with homology to the region of
CC TyrA which has chorismate mutase activity. The papB gene product is
CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
CC deoxy 4-aminoprephenate acid. Disruption of the papB and papC genes can
CC be used to produce strains of S.pristinaespiralis which are unable to
CC produce the antibiotic pristinamycin I but which may be able to produce
CC new, modified forms of it

XX SQ Sequence 387 BP; 50 A; 187 C; 110 G; 40 T; 0 U, 0 Other;

Alignment Scores:

Pred. No.:	1.03e-25	Length:	387
Score:	256.00	Matches:	51
Percent Similarity:	71.08%	Conservative:	8
Best Local Similarity:	61.45%	Mismatches:	24
Query Match:	49.81%	Indels:	0
DB:	2	Gaps:	0

US-10-089-514-4 (1-103) x AAT58552 (1-387)

OY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
DB 64 CTGACGCGCTGCGCGCGCTCGGACGCGCGGACGCGCCCTGTCGAGCCCGCCCGC 123
OY 27 ArgArgTleAspLeuGlyValArgTleAlaArgTyrSerArgHisGlyValProMet 46
DB 124 ACAAGCTGAGATGTGCTGCGCATCGCGGAGTACAGCGCCCTCCACGAGTCCGATG 183
OY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAspHis 66
DB 184 ATGACGCCCGACCGATGCGCCAGTCCAGTCCAGCGGACGCGCCGCTACGCCCGACCGC 243
OY 67 GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 86
DB 244 GGCATCGACCCCGCTTCTCGGACACCTGTACGACACGATCATCCAGACACTGCCCGC 303
OY 87 ValGluAsp 89
DB 304 CTCGAGGAC 312

RESULT 6
AAT59268 standard; cDNA; 2888 BP.

XX AC AAT59268;

DT 02-APR-1997 (first entry)

DE Streptomyces pristinaespiralis papA and papM intergenic region.

XX Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;

KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
 KM papC; isomerisation; aromatisation; N-methyltransferase; de.
 XX
 OS Streptomyces pristinaeaepristalis.
 FH
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT FT /*tag= a
 FT FT /product= "PapA"
 FT /note="C-terminal coding region only, i.e. a partial
 open reading frame"
 CDS complement(949..1836)
 FT FT /*tag= b
 FT FT /product= "PapC"
 FT 1873..2262
 FT FT /*tag= c
 FT FT /product= "PapB"
 FT 2259..2888
 FT FT /*tag= d
 FT FT /product= "PapM"
 FT /note="N-terminal coding region only, i.e. a partial
 open reading frame"
 FT
 XX
 XX MO9601901-A1.
 PN
 PD 25-JAN-1996.
 PF 04-JUL-1995; 95MO-FR000889.
 PR 08-JUL-1994; 94FR-00008478.
 XX
 XX (RHON) RHONE POULENC RORER SA.
 PA
 XX Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
 XX
 DR WPI; 1996-097631/10.
 DR P-PSDB; AAM11582, AAM11583.
 PT New streptogramin B derive. useful as antibiotics - produced by new
 PT mutants of Streptomyces having altered genes for streptogramin B
 PT biosynthesis.
 XX
 PS Example 1; Page 102-104; 146pp; French.
 XX
 CC The papA and papM genes of *S. pristinaeaepristalis* are involved in the
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
 CC pristinamycin 1A. The region between these two genes was sequenced and
 CC two open reading frames were identified. The first (papC) was on the
 CC complementary strand and decodes to an amino acid sequence with homology
 CC to the region of *E. coli* tyra which has been implicated in aromatisation
 CC reactions. The papC gene product is likely to be involved in a similar
 CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
 CC phenylpyruvate during DMPAPA synthesis. The second open reading frame
 CC (papB) could be decoded to give a product with homology to the region of
 CC tyra which has chorismate mutase activity. The papB gene product is
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
 CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
 CC used to produce strains of *S. pristinaeaepristalis* which are unable to
 CC produce the antibiotic pristinamycin I but which may be able to produce
 CC new, modified forms of it
 XX
 SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 1.63e-24 Length: 2888
 Score: 256.00 Matches: 51
 Percent Similarity: 71.08% Conservative: 8
 Best Local Similarity: 61.45% Mismatches: 24
 Query Match: 49.81% Indels: 0
 DB: 2 Gaps: 0
 US-10-089-514-4 (1-103) X AAT59268 (1-2888)

Qy 7 leuGlnArgLeuArgAlaGluAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
 Db 1936 CTCGACGCCCTGCGCCCTGCGCCCTGACCGCGGAGCGCCCTGCGACGCCCTCGC 1995
 Qy 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSerArgIleGlyValPromet 46
 Db 1996 ACAGCGCTGGAACATCGCGCATCGCGCATCGGAGATACAGAGCGCTCCACCAAGTCCGATG 2055
 Qy 47 MetGlnProGlyArgValSerLeuValIlyAspArgAlaAlaArgTyrAlaAlaAspHis 66
 Db 2056 ATCCACGCCCGCCGATCGCCGATCGCCGATCGCCGATCGCCGATCGCCGATCGCCGAT 2115
 Qy 67 GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 85
 Db 2116 GGCATCGACCCCGCTTCGCGCATCGCGCATCGCGCATCGCGCATCGCGCATCGCGCAT 2175
 Qy 87 ValGluAsp 89
 Db 2176 CTCGAGGAC 2184
 RESULT 7
 ACF71772
 ID ACF71772 standard; DNA; 360 BP.
 XX
 AC ACF71772;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Photorhabdus luminescens nucleotide sequence #10239.
 XX
 KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
 KW detection; food; gene expression; plant; animal; microorganism; toxin;
 KW antibiotic; bioplastic; virulence factor; disease model; plague;
 KW whooping cough; gene; ds.
 XX
 OS Photorhabdus luminescens.
 XX
 PN W0200294867-A2.
 XX
 PD 28-NOV-2002.
 XX
 PF 07-FEB-2002; 2002MO-IB003040.
 PR 07-FEB-2001; 2001FR-00001659.
 XX
 PA (INSP) INST PASTEUR.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 XX
 PI Duchaud E, Taourit S, Glaeier P, Frangeul L, Kunet F, Danchin A;
 PI Buchrieser C;
 XX
 DR WPI; 2003-148459/14.
 PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
 PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
 XX
 PS Claim 2; SEQ ID NO 10239; 1205bp; French.
 XX
 CC The invention relates to the isolation of genes and their encoded
 CC proteins from Photorhabdus luminescens. The isolated sequences are
 CC sources of probes and primers for detecting the genome of *P. luminescens*
 CC and related species; to study polymorphisms; for gene analysis and for
 CC detection/amplification of the genes. Antibodies (Ab) raised against the
 CC polypeptides encoded by the genes are used for detection/identification
 CC of *P. luminescens*, e.g. in foods. The genes, proteins, Ab and cells that
 CC carry a gene-containing vector are used to select compounds that
 CC modulate, regulate, induce or inhibit expression of the genes in plants,
 CC animals or microorganisms other than *P. luminescens* and are able to alter
 CC response or sensitivity to toxins and antibiotics produced by *P.*
 CC luminescens. Cells transformed to express the genes are useful for
 CC recombinant production of the proteins, particularly toxins and
 CC antibacterials useful as insecticides, bactericides and fungicides. The

CC genes/proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luumniscens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luumniscens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luumniscens genes

Alignment Scores:

Pred. No.:	4.9e-07	Length:	360
Score:	120.00	Matches:	25
Percent Similarity:	55.06%	Conservative:	24
Best local Similarity:	28.09%	Mismatches:	40
Query Match:	23.35%	Indels:	0
DB:	10	Gaps:	0

US-10-089-514-4 (1-103) X ACF71772 (1-360)

[illegible]

WP	ACF67367_49	Constitution (50 of 57) of ACF67367 from base 4900001 (Phototaphus luminescens nucleoti
WP	Sequence split into 57 fragments	Begin
WP	Fragment Name	End
WP	ACF67367_00	1 110000
WP	ACF67367_01	100001 210000
WP	ACF67367_02	200001 310000
WP	ACF67367_03	300001 410000
WP	ACF67367_04	400001 510000
WP	ACF67367_05	500001 610000
WP	ACF67367_06	600001 710000
WP	ACF67367_07	700001 810000
WP	ACF67367_08	800001 910000
WP	ACF67367_09	900001 1010000
WP	ACF67367_10	1000001 1110000
WP	ACF67367_11	1100001 1210000
WP	ACF67367_12	1200001 1310000
WP	ACF67367_13	1300001 1410000
WP	ACF67367_14	1400001 1510000
WP	ACF67367_15	1500001 1610000
WP	ACF67367_16	1600001 1710000
WP	ACF67367_17	1700001 1810000
WP	ACF67367_18	1800001 1910000
WP	ACF67367_19	1900001 2010000
WP	ACF67367_20	2000001 2110000
WP	ACF67367_21	2100001 2210000
WP	ACF67367_22	2200001 2310000
WP	ACF67367_23	2300001 2410000
WP	ACF67367_24	2400001 2510000
WP	ACF67367_25	2500001 2610000
WP	ACF67367_26	2600001 2710000
WP	ACF67367_27	2700001 2810000

WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5710000
WP	ACF67367_57	5700001	5810000
WP	ACF67367_58	5800001	5910000
WP	ACF67367_59	5900001	6010000
WP	ACF67367_60	6000001	6110000
WP	ACF67367_61	6100001	6210000
WP	ACF67367_62	6200001	6310000
WP	ACF67367_63	6300001	6410000
WP	ACF67367_64	6400001	6510000
WP	ACF67367_65	6500001	6610000
WP	ACF67367_66	6600001	6710000
WP	ACF67367_67	6700001	6810000
WP	ACF67367_68	6800001	6910000
WP	ACF67367_69	6900001	7010000
WP	ACF67367_70	7000001	7110000
WP	ACF67367_71	7100001	7210000
WP	ACF67367_72	7200001	7310000
WP	ACF67367_73	7300001	7410000
WP	ACF67367_74	7400001	7510000
WP	ACF67367_75	7500001	7610000
WP	ACF67367_76	7600001	7710000
WP	ACF67367_77	7700001	7810000
WP	ACF67367_78	7800001	7910000
WP	ACF67367_79	7900001	8010000
WP	ACF67367_80	8000001	8110000
WP	ACF67367_81	8100001	8210000
WP	ACF67367_82	8200001	8310000
WP	ACF67367_83	8300001	8410000
WP	ACF67367_84	8400001	8510000
WP	ACF67367_85	8500001	8610000
WP	ACF67367_86	8600001	8710000
WP	ACF67367_87	8700001	8810000
WP	ACF67367_88	8800001	8910000
WP	ACF67367_89	8900001	9010000
WP	ACF67367_90	9000001	9110000
WP	ACF67367_91	9100001	9210000
WP	ACF67367_92	9200001	9310000
WP	ACF67367_93	9300001	9410000
WP	ACF67367_94	9400001	9510000
WP	ACF67367_95	9500001	9610000
WP	ACF67367_96	9600001	9710000
WP	ACF67367_97	9700001	9810000
WP	ACF67367_98	9800001	9910000
WP	ACF67367_99	9900001	1000000

Alignment Scores:

Pred. No.:	0.00125	Length:	1100000
Score:	120.00	Matches:	25
Percent Similarity:	55.06%	Conservative:	24
Best Local Similarity:	28.09%	Mismatches:	40
Query Match:	23.35%	Indels:	0
DB:	10	Gaps:	0

US-10-089-514-4 (1-103) X ACF67367 49 (1-110000)

[illegible]

RESULT 9			
Continuation (2 of 7) of ACF65387 from base 100001 (Phototrhadus luminescens nucleotide			
ACF65387_1	Fragment Name	Begin	End
WP Sequence Split Into 7 Fragments	LOCUS ACF65387	Accession ACF65387	
WP	Fragment Name	Begin	End
WP	ACF65387_0	1	110000
WP	ACF65387_1	100001	210000
WP	ACF65387_2	200001	310000
WP	ACF65387_3	300001	410000
WP	ACF65387_4	400001	510000
WP	ACF65387_5	500001	610000
WP	ACF65387_6	600001	696798

Alignment Scores:

Pred. No.: 0.00125 Length: 110000
Score: 120.00 Matches: 25
Percent Similarity: 55.06% Conservative: 24
Best Local Similarity: 28.09% Mismatches: 40
Query Match: 23.35% Indels: 0
DB: 10 Gaps: 0

US-10-089-514-4 (1-103) x ACF65387_1 (1-110000)

QY 1 MetThrGluGlnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 13584 ATAAATGAACGGAATATTATACACTTACAGGAATCTTGATTATTAATATGACAG 13643
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSer 40
DB 13644 ATTTGGCACTGCTTCTGTAACGATGAAAGTTGTATGAAATCGCTGAGTAAAGCG 13703
QY 41 ArgHisGlyValPromeMetGlnProGlyArgValSerLeuValIleAspArgAlaAla 60
DB 13704 GGGCAGATATTCGAGATGATCCAGCCACACGTATTACCTGTTATGAAATATGTCGCG 13763
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
DB 13764 GATTAATCAACCGATTTGGGTTACGGCTGAATATATCTGAATCAATATTCAACTGCTT 13823
QY 81 IleThrGluMetCysArgValGluAsp 89
DB 13824 ATTGAAGAGACTTGCCGTCGAGAA 13850

RESULT 10

Continuation (13 of 17) of AAV21209 from base 120001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession AAV21209

WP	Fragment Name	Begin	End
WP	AAV21209_00	1	110000
WP	AAV21209_01	100001	210000
WP	AAV21209_02	200001	310000
WP	AAV21209_03	300001	410000
WP	AAV21209_04	400001	510000
WP	AAV21209_05	500001	610000
WP	AAV21209_06	600001	710000
WP	AAV21209_07	700001	810000
WP	AAV21209_08	800001	910000
WP	AAV21209_09	900001	1010000
WP	AAV21209_10	1000001	1110000
WP	AAV21209_11	1100001	1210000
WP	AAV21209_12	1200001	1310000
WP	AAV21209_13	1300001	1410000
WP	AAV21209_14	1400001	1510000
WP	AAV21209_15	1500001	1610000
WP	AAV21209_16	1600001	1664976

Alignment Scores:

Pred. No.: 0.0158 Length: 110000
Score: 112.00 Matches: 20
Percent Similarity: 53.09% Conservative: 23
Best Local Similarity: 24.69% Mismatches: 38
Query Match: 21.79% Indels: 0
DB: 2 Gaps: 0

US-10-089-514-4 (1-103) x AAV21209_12 (1-110000)

QY 1 MetThrGluGlnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 38358 GTTACATGATGATAGAACTGCTGAATATGAGAAAGATTGATGACAAATAG 38417
QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSer 40
DB 38418 ATATTTAAGCTAATTTGCTGAAGAAATAGTTTACCTAAGATGATGAGATTAATAAT 38477
QY 41 ArgHisGlyValPromeMetGlnProGlyArgValSerLeuValIleAspArgAlaAla 60

DB 38478 CAGCTTGATTCCTATTAAAGACCAGAAAGAAAATATATATATACGATAGATATAG 38537

QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80

DB 38538 AAACCTTTGAAGAACATACGTTGATGAAATATTTGCATTAATAATTTCAAAATCTT 38597

QY 81 Ile 81

DB 38598 ATA 38600

RESULT 11

Continuation (17 of 17) of ADM27081 from base 160001 (Hypethermophile Methanopyrus ka
WP Sequence split into 17 fragments LOCUS ADM27081 Accession ADM27081

WP	Fragment Name	Begin	End
WP	ADM27081_00	1	110000
WP	ADM27081_01	100001	210000
WP	ADM27081_02	200001	310000
WP	ADM27081_03	300001	410000
WP	ADM27081_04	400001	510000
WP	ADM27081_05	500001	610000
WP	ADM27081_06	600001	710000
WP	ADM27081_07	700001	810000
WP	ADM27081_08	800001	910000
WP	ADM27081_09	900001	1010000
WP	ADM27081_10	1000001	1110000
WP	ADM27081_11	1100001	1210000
WP	ADM27081_12	1200001	1310000
WP	ADM27081_13	1300001	1410000
WP	ADM27081_14	1400001	1510000
WP	ADM27081_15	1500001	1610000
WP	ADM27081_16	1600001	1694968

Alignment Scores:

Pred. No.: 0.0333 Length: 94968
Score: 109.00 Matches: 25
Percent Similarity: 53.01% Conservative: 19
Best Local Similarity: 30.12% Mismatches: 35
Query Match: 21.21% Indels: 4
DB: 1 Gaps: 1

US-10-089-514-4 (1-103) x ADM27081_16 (1-94968)

QY 5 ArgGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThr 24
DB 17909 AACGTGCTCGAGAGCTCAAGCGCGAATATGACCGCATGATGATGTTGCTGACGCA 17850
QY 25 ValArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIleSerArgHisGlyVal 44
DB 17849 GTCATCGAGCGCTCGAAGGTGCGCGGGAATATGGCGCTCAAGAGGGGCTTA 17790
QY 45 PromeMetGlnProGlyArgValSerLeuValIleAspArgAlaAlaArgTyrAlaAla 64
DB 17789 CCACTGACGAGCAGAGGCGGAGAGAGAGCTCCGGAGAGTGCAGAGAGGTTTAA 17730
QY 65 AspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMet 84
DB 17729 ACAGAGGGGCTGATCGCGG-----CTCGCGGACATCGTGTGCTTCCTATC 17682
QY 85 CysArgVal 87
DB 17681 CTAAAGCTT 17673
RESULT 12
ADP00853
ID ADP00853 standard; DNA; 1125 BP.
AC ADP00853;
XX
XX 12-FEB-2004 (first entry)
XX
XX Bacterial polynucleotide #1138.
XX

KM	Proteus mirabilis infection; bacterial infection; antibacterial;
KM	Immunostimulant; gene; ds.
XX	
OS	Proteus mirabilis.
PN	US6605709-B1.
XX	
PD	12-AUG-2003.
XX	
PF	05-APR-2000; 2000US-00543681.
XX	
PR	09-APR-1999; 99US-0128706P.
XX	
PA	(GENO-) GENOME THERAPEUTICS CORP.
XX	
PI	Bretton GL;
XX	
DR	WPI: 2003-895291/82.
DR	P-PSDB; ADF05025.
XX	
PT	New Proteus mirabilis polypeptides and polynucleotides, useful as
PT	reagents for diagnosis of bacterial disease, as components of
PT	antibacterial vaccines, as targets for antibacterial drugs, or as
PT	biocontrol agents for plants.
XX	
PS	Disclosure; SEQ ID NO 1138; 870pp; English.
XX	
CC	The invention relates to new Proteus mirabilis polypeptides and
CC	polynucleotides. The invention also relates to antibodies against the
CC	polypeptides, methods for producing the polypeptides, a method of
CC	generating vaccines for immunising an individual against P. mirabilis, a
CC	method for evaluating a compound for the ability to bind a P. mirabilis
CC	polypeptide and a method for screening test compounds for anti-bacterial
CC	activity. The polypeptides and polynucleotides are useful as molecular
CC	targets for diagnosing, preventing and treating pathological conditions
CC	resulting from bacterial infection, as reagents for diagnosis of
CC	bacterial diseases, as components of antibacterial vaccines, as targets
CC	for antibacterial drugs or as bio-control agents for plants. This
CC	sequence represents a Proteus mirabilis polynucleotide of the invention.
XX	
SQ	Sequence 1125 BP; 352 A; 180 C; 250 G; 343 T; 0 U; 0 Other;
XX	
Alignment Scores:	
Pred. No.:	0.000197 Length: 1125
Score:	106.00 Matches: 27
Percent Similarity:	46.94% Conservative: 19
Best Local Similarity:	27.55% Mismatches: 36
Query Match:	20.62% Indels: 16
DB:	10 Gaps: 2
US-10-089-514-4 (1-103) x ADF00853 (1-1125)	
QY	6 GluUeuGlnArgLeuArgAlaGluUeuAAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
DB	10 GAATTTACAAAATTTCGAGAGCAAAATGATTCAGGGGATAAACCACTTCCTTACTC 69
QY	26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
DB	70 GCCAAAAGATTCGAATTCGTAGCAGAAAGTGGGAGAGATTAAAAATCCGCACGGTTACT 122
QY	46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
DB	130 ATCTATGCACACATAGAGAGCGCAATGCTTGCTTCGACGTAATGAAGCAGAAAAA 169
QY	66 HisGlyLeuAspGluSerPheLeuValaLeuLeuTyrAspValIleIleThrGluMetCys 85
DB	190 ATGGATTTCTCTGATTTA-----MetSerArgGluSerLeuThrAlaGluAsp 101
QY	86 ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101
DB	211 ---ATTGAAGATGTTTACGTCGAGTAATGAAGAGAGTCATATAGCAAAAGAAAT 261
RESULT 13	

[illegible]

Best Local Similarity: 27.55% Mismatches: 33
 Query Match: 19.84% Indels: 16
 DB: 12 Gaps: 2

US-10-089-514-4 (1-103) x ADJ98178 (1-1122)

```

QY      6  GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB     10  GAACTGACCCGCGTTACCGCATCAAAATTGACAGTAGATTAAGCCGCTGCGATCTGCTG 69
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY     26  ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIysSerArgHisGlyValPro 45
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB     70  GCTAAGCGACTGGAACCTGCTGCGCCGAGGTAGGTGAGGTGAAGAGCCGTTACGGCTGCGCT 129
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY     46  MetMetGlnProGlyArgValSerLeuValIysAspArgAlaAlaArgTyrAlaAlaAsp 65
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB    130  ATCTATGTGCTGCTGAGCGTGAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 189
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY     66  HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB    190  CTCGGCGTACACCGGATCTG----- 210
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
QY     86  ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB    211  --ATTGAGGATGTGCTGCTGCGGTGATCGGGAATCTTATACCAAGCGAAT 261
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
  
```

Search completed: October 5, 2005, 10:49:05
 Job time : 243.876 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 10:07:08 : Search time 56.3672 Seconds
(without alignments)
2989.973 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514
Sequence: 1 MTEQNELQLRAELDALDGT.....MCRVEDLVMSRESLTAEDRR 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 81813359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+g2n.model -DEV=xlh
-Q/cgn2_1/USPTO.spool_h/US10089514/runat_04102005_105746_8071/app_query.fasta_1.1621
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=bl0sum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514_@CGN_1_116_@runat_04102005_105746_8071 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/prodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/prodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/prodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/prodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/prodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	256	49.8	387	3	US-08-765-907A-4
2	256	49.8	387	3	US-09-987-614A-4
3	256	49.8	2888	3	US-08-765-907A-1
4	256	49.8	2888	4	US-09-987-614A-1
5	112	21.8	1664976	4	US-08-916-421B-1
6	112	21.8	1664976	4	US-09-692-570-1
7	106	20.6	1135	4	US-09-543-681A-1138
8	105	20.4	1305	4	US-09-489-039A-4503
9	104.5	20.3	1963	4	US-09-902-540-2740
10	104.5	20.3	15689	4	US-09-902-540-1129
11	98	19.1	850	4	US-09-221-017B-361
12	93	18.1	1131	4	US-09-489-039A-4572

13	90	17.5	1038	4	US-09-902-540-5637	Sequence 5637, Ap
14	90	17.5	49225	4	US-09-902-540-1269	Sequence 1269, Ap
15	88.5	17.2	804	4	US-09-252-991A-4814	Sequence 4814, Ap
16	88.5	17.2	960	4	US-09-252-991A-4705	Sequence 4705, Ap
17	88.5	17.2	1653	4	US-09-252-991A-4724	Sequence 4724, Ap
18	88.5	17.2	1791	4	US-09-252-991A-4684	Sequence 4684, Ap
19	88.5	17.2	1830121	4	US-09-557-884-1	Sequence 1, Appli
20	88.5	17.2	1830121	4	US-09-643-990A-1	Sequence 1, Appli
21	82	16.0	1167	4	US-09-543-681A-1120	Sequence 1320, Ap
22	76	14.8	267	4	US-09-583-110-1836	Sequence 1836, Ap
23	75	14.6	4258	3	US-07-765-830A-5	Sequence 5, Appli
24	75	14.6	4825	4	US-09-949-016-15912	Sequence 15912, A
25	74	14.4	888	4	US-09-252-991A-5830	Sequence 5830, Ap
26	74	14.4	1041	4	US-09-252-991A-5815	Sequence 5815, Ap
27	74	14.4	1066	4	US-09-252-991A-10259	Sequence 10259, A
28	74	14.4	1101	4	US-09-252-991A-10597	Sequence 10597, A
29	74	14.4	1581	4	US-09-252-991A-10473	Sequence 10473, A
30	74	14.4	1611	4	US-09-252-991A-5778	Sequence 5778, Ap
31	74	14.4	9839	4	US-09-902-540-996	Sequence 996, App
32	73	14.2	921	4	US-09-902-540-7847	Sequence 7847, Ap
33	73	14.2	5125	4	US-09-902-540-785	Sequence 785, App
34	72	14.0	447	4	US-09-902-540-9336	Sequence 9326, Ap
35	71.5	13.9	1230	4	US-10-105-319-1	Sequence 1, Appli
36	71.5	13.9	1281	3	US-09-105-537-19	Sequence 19, Appli
37	71.5	13.9	13613	3	US-09-105-537-3	Sequence 3, Appli
38	71.5	13.9	18431	3	US-09-320-878-19	Sequence 1090, Ap
39	71.5	13.9	38506	3	US-09-141-908-1	Sequence 19, Appli
40	71.5	13.9	38506	4	US-09-657-440-19	Sequence 19, Appli
41	71.5	13.9	38506	4	US-09-902-540-7256	Sequence 7256, Ap
42	71	13.8	6744	4	US-09-902-540-682	Sequence 682, App
43	71	13.8	1524	4	US-09-540-236-1471	Sequence 1471, Ap
44	69.5	13.5	1524	4	US-09-902-540-8770	Sequence 8770, Ap
45	69	13.4	1125	4		

ALIGNMENTS

RESULT 1
US-08-765-907A-4
; Sequence 4, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BMMAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: PAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; TITLE OF INVENTION: Metasyntesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765, 907A
; CURRENT FILING DATE: 1997-03-20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
; US-08-765-907A-4

Alignment Scores:

Pred. No.: 4,96e-28
Score: 256.00
Percent Similarity: 71.08%
Best Local Similarity: 61.45%
Query Match: 49.81%
DB: 3
Length: 387
Matches: 51
Conservative: 8
Mismatch: 24
Indels: 0
Gaps: 0


```
/ APPLICANT: THIBAUT, Denis
/ APPLICANT: BAWAS-JACQUES, Nathalie
/ APPLICANT: BLANCHE, Francis
/ APPLICANT: COUZET, Joel
/ APPLICANT: BARRIERE, Jean-Claude
/ APPLICANT: DEBUSCHÉ, Laurent
/ APPLICANT: FAMECHON, Alain
/ APPLICANT: PARIS, Jean-Marc
/ APPLICANT: DUTRUC-ROSSER, Gilles
/ TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
/ FILE REFERENCE: Streptogramin genes
/ CURRENT FILING DATE: 2001-11-15
/ PRIOR APPLICATION NUMBER: US/09/987,614A
/ PRIOR FILING DATE: 1997-03-20
/ NUMBER OF SEQ ID NOS: 17
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 1
/ LENGTH: 2888
/ TYPE: DNA
/ ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

Alignment Scores:
Pred. No.:      8,62e-27      Length:      2888
Score:          256.00      Matches:      51
Percent Similarity: 71.08%      Conservative: 8
Best Local Similarity: 61.45%      Mismatches: 24
Query Match:    49.81%      Indels:      0
DB:              4          Gaps:      0

US-10-089-514-4 (1-103) x US-09-987-614A-1 (1-2888)

QY      7      LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
Db      1936  CTCGACCGCTGCGCGCGCGCGCTGAGCGCGCGCGCTGCGAGCGCGCTGCGAGCGCGCTCGC 1995

QY      27      ArgArgGLeuAspLeuGlyValArgGLeuAlaArgTyrTyrSerArgHisGlyValArgPromet 46
Db      1996  ACACGCGCTGACATCTCGCTCGGATCGGCGAGTACAGCGCGCTCCACAGTGCATG 2055

QY      47      MetGlnProGlyArgValSerLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 66
Db      2056  ATGACGCCCAACCGGATCGCCAGGTCCACGCCCAACGCCCGCTCCAGCCGCGACAC 2115

QY      67      GlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCysArg 86
Db      2116  GGCATCGACCCCGCTTCTGCGCAGCCCTGTACAGACGATCATCACCGAGACTGCGCGC 2175

QY      87      ValGluAsp 89
Db      2176  CTCGAGGAC 2184

RESULT 5
US-08-916-421B-1
/ Sequence 1, Application US/08916421B
/ Patent No. 6503729
/ GENERAL INFORMATION:
/ APPLICANT: Built et al.
/ TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
/ Patent No. 6503729
/ TITLE OF INVENTION: Jannaschii
/ FILE REFERENCE: PB275
/ CURRENT APPLICATION NUMBER: US/08/916,421B
/ CURRENT FILING DATE: 1997-08-22
/ PRIOR APPLICATION NUMBER: US 60/024,428
/ PRIOR FILING DATE: 1996-08-22
/ NUMBER OF SEQ ID NOS: 3
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 1664976
/ TYPE: DNA
```

```
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (10398)..(10398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (319226)..(319226)
```

```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491)..(1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091)..(1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020)..(1569020)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854)..(1664854)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Alignment Scores:
Pred. No.: 0.0612 Length: 1664976
Score: 112.00 Matches: 20
Percent Similarity: 53.09% Conservative: 23
Best Local Similarity: 24.69% Mismatches: 38
Query Match: 21.79% Indels: 0
DB: 4 Gaps: 0

US-10-089-514-4 (1-103) x US-08-916-421B-1 (1-1664976)

QY 1 MetThGlUGlnsnglUleuglnArgLeuArgAlaGlUleuaspAlaleuaspGlyThr 20
DB 1238358 GTTACATGATAGAGAACTTCCTGAAATTAGGAAAGATGATGACATTAAG 1238417
QY 21 LeuLeuaspThrValArgArgArgIleAspLeuGlyValArgIleAlaArgIleYsser 40
DB 1238418 ATATTAAAGCTTAATGCTGAAAGAAATTAATTAGCTTAAGCATGTGCTGATTAATAAT 1238477
QY 41 ArgHleGlyValProMetMetcInProGlyArgValSerLeuValLysAspArgAlaIa 60
DB 1238478 CAGCTTGATTCCTATTTAACGACCCAGAAAGAAAATATATATACATGATTAAGA 1238537
QY 61 ArgTyrAlaAlaAspHisGlyLeuaspGlySerPheLeuValAsnLeuTyrAspValIle 80
DB 1238538 AAACCTTGTAAGAAACATTAACGTGATGAAATATGCGATTAAATATTCATAATACCTT 1238597
QY 81 Ile 81
DB 1238598 ATA 1238600

RESULT 6
US-09-692-570-1
; Sequence 1, Application US/09692570
; Patent No. 6797466
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococ
; Patent No. 6797466
; TITLE OF INVENTION: jannaschi
; FILE REFERENCE: PB275C1
; CURRENT APPLICATION NUMBER: US/09/692,570
; PRIOR FILING DATE: 2003-01-14
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; PRIOR APPLICATION NUMBER: US 08/916,421
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschi
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28232)..(28232)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
```

OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (103998)..(103998)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (231980)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g

FEATURE:
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:

```
/ NAME/KEY: misc_feature
/ LOCATION: (1084830)..(1084830)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1095846)..(1095846)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1119881)..(1119881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1130881)..(1130881)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1310988)..(1310988)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1313224)..(1313224)
/ OTHER INFORMATION: n equals a, t, c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1349473)..(1349473)
/ OTHER INFORMATION: n equals a, t, c, or g
```

Alignment Scores:

Pred. No.:	0.0612	Length:	1664976
Score:	112.00	Matches:	20
Percent Similarity:	53.09%	Conservative:	23
Best Local Similarity:	24.69%	Mismatches:	38
Query Match:	21.79%	Indels:	0
DB:	4	Gaps:	0

US-10-089-514-4 (1-103) x US-09-692-570-1 (1-1664976)

```
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1238358 GTTACAGATGATAGAGAACTTGGTGAATTAGAGAGAGATTGATGATTCGACATTAAG 1238417
QY 21 LeuLeuAspThrValAlaArgArgGlyLeaAspLeuGlyValAlaArgIleAlaArgIlySer 40
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1238418 ATATTAAAGCTAATCTCTGAAAGAAATAGTTTACTAGAGATGCTGAGATMAAAAT 1238477
QY 41 ArgHisGlyValAlaProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAla 60
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1238478 CAGCTTGATTTCTCTATTAAAGACCCAGAAAGAAAAATATATATACGATTAGAAATAGA 1238537
QY 61 ArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIle 80
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 1238538 AATCTTTGTAAAGACATTACGTTGATGAAAAATATTGGCATTAATAATTTCAAAATACT 1238597
QY 81 Ile 81
    |||
Db 1238598 ATA 1238600

RESULT 7
US-09-543-681A-1138
/ Sequence 1138, Application US/09543681A
/ Patent No. 6605709
/ GENERAL INFORMATION:
/ APPLICANT: GARY BRETON
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
/ FILE REFERENCE: 2709.1002-001
/ CURRENT APPLICATION NUMBER: US/09/543,681A
/ PRIOR FILING DATE: 2000-04-05
/ PRIOR APPLICATION NUMBER: US 60/128,706
/ PRIOR FILING DATE: 1999-04-09
/ NUMBER OF SEQ ID NOS: 8344
/ SEQ ID NO 1138
```

```
/ LENGTH: 1125
/ TYPE: DNA
/ ORGANISM: Proteus mirabilis
US-09-543-681A-1138
```

Alignment Scores:

Pred. No.:	1.43e-05	Length:	1125
Score:	106.00	Matches:	27
Percent Similarity:	46.94%	Conservative:	19
Best Local Similarity:	27.55%	Mismatches:	36
Query Match:	20.62%	Indels:	16
DB:	4	Gaps:	2

US-10-089-514-4 (1-103) x US-09-543-681A-1138 (1-1125)

```
QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 10 GAATTACAAATTTGGAGACCAATTGATCGAGTGAGTAAACATTGCTTCTTACTC 69
QY 26 ArgArgArgIleAspLeuGlyValAlaArgIleAlaArgIlySerArgHisGlyValPro 45
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 70 GCCAAAGAAATGCAATTTGTACAGAGAGTGGAGAAAGTTAAAAATCGCCAGCGTTACT 129
QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 ATCTATGACACAGATAGAGAACCCGCAATGCTTCTTCGACGTAAATGACAGAAAAA 189
QY 66 HisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85
    |||||
Db 190 ATGGGATTTTCGTCGATTTA----- 210
QY 86 ArgValGluAspLeuVal-----MetSerArgGluSerLeuThrAlaGluAsp 101
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 211 ---ATTGAAGATGTTTACTGCGAGTAATAGAGAGTCTATATAGCAAGAAAT 261
```

RESULT 8

US-09-489-039A-4503
/ Sequence 4503, Application US/09489039A
/ Patent No. 6610836
/ GENERAL INFORMATION:

APPLICANT: GARY BRETON et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
FILE REFERENCE: 2709.2004001
CURRENT APPLICATION NUMBER: US/09/489,039A
PRIOR FILING DATE: 2000-01-27
PRIOR APPLICATION NUMBER: US 60/117,747
NUMBER OF SEQ ID NOS: 14342
SEQ ID NO 4503
LENGTH: 1305
TYPE: DNA
ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4503

Alignment Scores:

Pred. No.:	2.46e-05	Length:	1305
Score:	105.00	Matches:	29
Percent Similarity:	48.45%	Conservative:	18
Best Local Similarity:	29.90%	Mismatches:	44
Query Match:	20.43%	Indels:	6
DB:	4	Gaps:	1

US-10-089-514-4 (1-103) x US-09-489-039A-4503 (1-1305)

```
QY 1 MetThrGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 145 ATGACCGAGAGAAACCATTAATCTGGCGCTGCCGATTAAGATAGCGCCCTGAGACGAGAAG 204
QY 21 LeuLeuAspThrValAlaArgArgGlyLeaAspLeuGlyValAlaArgIleAlaArgIlySer 40
    |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 205 CTGCTTGCTCTGCTGCGAGCGCGCGGACTGCTGTGAGCTTGAAGAAAGCCAAACTG 264
```



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Montoy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 361:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 850 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEetical: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1...850
; US-09-221-017B-361

Alignment Scores:
Pred. No.: 0.00014 Length: 850
Score: 98.00 Matches: 21
Percent Similarity: 56.34% Conservative: 19
Best Local Similarity: 29.58% Mismatches: 31
Query Match: 19.07% Indels: 0
DB: 3 Gaps: 0

US-10-089-514-4 (1-103) x US-09-221-017B-361 (1-850)

QY 3 GluClasngluenuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeu 22
Db 214 CAAGACGACGAGCTGATCTTGGCGCATCGCAAGATTGATGAGATGAGATATAGTG 155
QY 23 AspThrValArgArgArgLeuAspLeuGlyValArgIleAlaArgTyrLysSerArgHis 42
Db 154 GAATTGCTAGCTGCGCGCATGCAAGTGCATACGAGTACGTTCTTCAAAAAAGACAC 95
QY 43 GlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyr 62
Db 94 AATCTGCTGTGGTTCGATCTCCGCTACCAACAACACTACAGCGCAACCGTCCCGTACT 35
QY 63 AlaAlaAspHisGlyLeuAspGlyUserPheLeu 73
Db 34 GCAGCCCTTGTAGGTTTGACGAAACATTATATA 2

RESULT 12
US-09-489-039A-4572
; Sequence 4572, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4572
; LENGTH: 1131
; TYPE: DNA
```

```

; ORGANISM: Klebsiella pneumoniae
; US-09-489-039A-4572

Alignment Scores:
Pred. No.: 0.00112 Length: 1131
Score: 93.00 Matches: 26
Percent Similarity: 47.96% Conservative: 21
Best Local Similarity: 26.53% Mismatches: 35
Query Match: 18.09% Indels: 16
DB: 4 Gaps: 2

US-10-089-514-4 (1-103) x US-09-489-039A-4572 (1-1131)

QY 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
Db 19 GAAGTACCGCGGCTTACGCATCAATGTATGAAAGTGAATGAAGCGTTGCTTACCGCGTGC 78
QY 26 ArgArgArgTLeuAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
Db 79 GCTAACCGTCTGAGCTGCTGCGCCGCAAGTCCGAGTGAAGACCATATGCGCTGCGCG 138
QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
Db 139 ATTACGTCCTCCGAGCGCAGATCGGATGCTCCCTCCGACGTGAAGAGCGGCTGCG 198
QY 66 HisGlyLeuAspGlyUserPheLeuValAsnLeuTyrAspValIleIleThrGluMetCys 85
Db 199 CTCGGCGTACCAACCGGATCTG----- 219
QY 86 ArgValGluAspLeuVal-----MetSerArgLysSerLeuThrAlaGluAsp 101
Db 220 ---ATCGAAGACGTTCTGCGTCGCGTGAATCGGATCTTCCATTCGCGCAAAAC 270

RESULT 13
US-09-902-540-5637
; Sequence 5637, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; CURRENT FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 5637
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
; US-09-902-540-5637

Alignment Scores:
Pred. No.: 0.0027 Length: 1038
Score: 90.00 Matches: 20
Percent Similarity: 53.33% Conservative: 20
Best Local Similarity: 26.67% Mismatches: 35
Query Match: 17.51% Indels: 0
DB: 4 Gaps: 0

US-10-089-514-4 (1-103) x US-09-902-540-5637 (1-1038)

QY 7 LeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
Db 775 CTGACCGAAGTTCGGAGTGCCTGATGCGTTGACCCGACCTGTTCAACTCCCTGAAT 834
QY 27 ArgArgGlyLeuAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMet 46
Db 835 CGTGGCGCCCACTCATCCAGAGCGCGCACCTGAAGGCGGACGAGCATTCGCTGCGTGC 894
```

```

QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66
DB 895 CCAGACCGCGGAGCGCGAGCGAGCTCTGTGAGACCCGGCGCTGCGCGGAGACG 954
QY 67 GlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 955 GGCATGACCGCGGAGCGAGACGAGAAAGCTGTTCGCGCCGCTCTG 999

RESULT 14
US-09-902-540-1269/C
; Sequence 1269, Application US/09902540
; Patent No. 6833447
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO 1269
; LENGTH: 49225
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1269

Alignment Scores:
Pred. No.: 0 648 Length: 49225
Score: 90.00 Matches: 20
Percent Similarity: 53.33% Conservative: 20
Best Local Similarity: 26.67% Mismatches: 35
Query Match: 17.51% Indels: 0
DB: Gaps: 0

US-10-089-514-4 (1-103) x US-09-902-540-1269 (1-49225)
QY 7 LeuGlnArgLeuArgAlaGlnLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
DB 40904 CTGACCGAGAGTTCGCGAGCTGCGATGCGGTGAGCCGTCGATCACTCTGAAT 40845
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValProMet 46
DB 40844 CGTCGCGCCGACCTCATCCAGCGCCGACACCTGAAAGCGCGAGCAGGACTTCCGCTG 40785
QY 47 MetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAspHis 66
DB 40784 CCAGACCGCGGAGCGCGAGCGAGCTCTGTGAGACCCGGCGCTGAGCCGCGAGACG 40725
QY 67 GlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 40724 GGCATGACCGCGGAGCGAGACGAGAAAGCTGTTCGCGCCGCTCTG 40680

RESULT 15
US-09-252-991A-4814/C
; Sequence 4814, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 31142
; SEQ ID NO 4814

```

```

; LENGTH: 804
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-4814

Alignment Scores:
Pred. No.: 0 0031 Length: 804
Score: 88.50 Matches: 24
Percent Similarity: 53.95% Conservative: 17
Best Local Similarity: 31.58% Mismatches: 32
Query Match: 17.22% Indels: 3
DB: Gaps: 2

US-10-089-514-4 (1-103) x US-09-252-991A-4814 (1-804)
QY 7 LeuGlnArgLeuArgAlaGlnLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg 26
DB 690 CTGCGGAGACTTCGCGAGCGGACATCGAGATGACTTGATTCCTCCAGGCCCTCGGC 631
QY 27 ArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
DB 630 CGCGGATGAGACTCACTCAAGCGCGGCTGCGGCTTCAAGCCAGCGGCGGATTCG 571
QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
DB 570 -----GCACCGGAGGCGGCTGCGGAGATGCTCCCGAGCGCGCGCTGCGCGAGAA 517
QY 66 HisGlyLeuAspGlnSerPheLeuValAsnLeuTyrAspValIleIle 81
DB 516 AACGAGCTGACCGCGGCTTGTGAGGAGACTGTTCGCGGATCATC 469

Search completed: October 5, 2005, 21:11:55
Job time : 605.367 secs

```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 226.674 Seconds

(without alignments)
3137.453 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514

Sequence: 1 MTEQNELQRRARLALDGT.....MCRVEDLVMSRESLTAEDDR 103

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:

Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-O=/cg2_1/USPTO.spool.h/US10089514/runat_04102005_105745_8055/app_query.fasta_1.1621
-DB=Published Applications NA -QFMT=fastab -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=1 -MATRIX=blonum62
-TRANS=human40.cdt -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10089514@cgn_1.1.920@runat_04102005_105745_8055
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications_NA:*

1: /cg2_6/prodata/1/pubna/US07_PUBCOMB.seq:*
2: /cg2_6/prodata/1/pubna/PCT_NEW_PUB.seq:*
3: /cg2_6/prodata/1/pubna/US06_NEW_PUB.seq:*
4: /cg2_6/prodata/1/pubna/US06_PUBCOMB.seq:*
5: /cg2_6/prodata/1/pubna/US07_NEW_PUB.seq:*
6: /cg2_6/prodata/1/pubna/PCTUS_PUBCOMB.seq:*
7: /cg2_6/prodata/1/pubna/US08_NEW_PUB.seq:*
8: /cg2_6/prodata/1/pubna/US08_PUBCOMB.seq:*
9: /cg2_6/prodata/1/pubna/US09_PUBCOMB.seq:*
10: /cg2_6/prodata/1/pubna/US09B_PUBCOMB.seq:*
11: /cg2_6/prodata/1/pubna/US09C_PUBCOMB.seq:*
12: /cg2_6/prodata/1/pubna/US09_NEW_PUB.seq:*
13: /cg2_6/prodata/1/pubna/US10A_PUBCOMB.seq:*
14: /cg2_6/prodata/1/pubna/US10B_PUBCOMB.seq:*
15: /cg2_6/prodata/1/pubna/US10C_PUBCOMB.seq:*
16: /cg2_6/prodata/1/pubna/US10D_PUBCOMB.seq:*
17: /cg2_6/prodata/1/pubna/US10E_PUBCOMB.seq:*
18: /cg2_6/prodata/1/pubna/US10F_PUBCOMB.seq:*
19: /cg2_6/prodata/1/pubna/US10G_PUBCOMB.seq:*
20: /cg2_6/prodata/1/pubna/US10H_PUBCOMB.seq:*
21: /cg2_6/prodata/1/pubna/US10I_PUBCOMB.seq:*
22: /cg2_6/prodata/1/pubna/US10J_PUBCOMB.seq:*
23: /cg2_6/prodata/1/pubna/US11_PUBCOMB.seq:*
24: /cg2_6/prodata/1/pubna/US11_NEW_PUB.seq:*
25: /cg2_6/prodata/1/pubna/US60_NEW_PUB.seq:*
26: /cg2_6/prodata/1/pubna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB ID	Description
1	514	100.0	312	20	US-10-472-587-3
2	486.5	94.6	3305	14	US-10-126-927-68
3	439	85.4	12391	14	US-10-126-927-67
4	102	19.8	1122	18	US-10-137-310-1
5	102	19.8	1122	18	US-10-634-548-16
6	102	19.8	1152	19	US-10-433-556-20
7	98	19.1	860	13	US-10-194-163-361
8	96.5	18.8	1833	17	US-10-381-779-40
9	91	17.7	387	15	US-10-156-761-6187
10	91	17.7	9025608	15	US-10-156-761-1
11	90	17.5	1132	16	US-10-137-310-3
12	90	17.5	1132	19	US-10-433-556-18
13	90	17.5	1238	16	US-10-380-132-7
14	89.5	17.4	1104	17	US-10-282-122A-15346
15	88.5	17.2	1830121	17	US-10-329-670-1
16	88.5	17.2	1830121	20	US-10-158-865-1
17	88.5	17.2	1830121	22	US-10-981-687-1
18	86	16.7	951	15	US-10-156-761-2533
19	82.5	16.1	300	18	US-10-335-977-4562
20	82.5	16.1	303	18	US-10-335-977-4561
21	78.5	15.3	1868	13	US-10-027-632-99311
22	78.5	15.3	1868	13	US-10-027-632-99312
23	78.5	15.3	1868	17	US-10-027-632-99311
24	78.5	15.3	1868	17	US-10-027-632-99312
25	78.5	15.3	1869	13	US-10-027-632-98019
26	77.5	15.1	1869	17	US-10-027-632-98019
27	77.5	15.0	1920	22	US-10-450-763-30061
28	77	15.0	1920	22	US-10-472-928-2611
29	75	14.6	821	13	US-10-027-632-166491
30	75	14.6	821	17	US-10-027-632-166491
31	75	14.6	1111	17	US-10-369-493-24241
32	75	14.6	1699	15	US-10-102-528-127
33	75	14.6	216298	21	US-10-472-928-4979
34	75	14.4	2282	17	US-10-369-493-36322
35	74	14.4	729	15	US-10-156-761-3396
36	73	14.2	1017	15	US-10-146-772-249
37	73	14.2	1017	17	US-10-241-742-249
38	73	14.2	1017	17	US-10-440-523-249
39	73	14.2	1017	18	US-10-440-503-249
40	73	14.2	1017	18	US-10-461-925-249
41	73	14.2	1882	17	US-10-320-797-2260
42	73	14.2	2089	17	US-10-320-797-1260
43	73	14.2	4089	17	US-10-320-797-260
44	73	14.2	2089	17	US-10-320-797-260
45	73	14.2	9025608	15	US-10-156-761-1

RESULT 1
US-10-472-587-3
; Sequence 3, Application US/10472587
; Publication No. US20040214274A1
; GENERAL INFORMATION:
; APPLICANT: YANAI, Naomi
; APPLICANT: YANAI, Koji
; APPLICANT: WATANABE, Manabu
; APPLICANT: MORIYA, Tatsuki
; APPLICANT: MURAKAMI, Takehiko
; TITLE OF INVENTION: Transformants Producing Substance P1022 Derivatives, Methods for
; FILE REFERENCE: 2003-1302A/MMC/00144
; CURRENT APPLICATION NUMBER: US/10/472,587
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 82227/2001

ALIGNMENTS

```
/ PRIOR FILING DATE: 2001-03-22
/ NUMBER OF SEQ ID NOS: 40
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 3
/ LENGTH: 312
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
/ FEATURE:
/ NAME/KEY: CDS
/ LOCATION: (1)..(309)
US-10-472-587-3

Alignment Scores:
Pred. No.: 1,266-68 Length: 312
Score: 514.00 Matches: 103
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 20 Gaps: 0

US-10-089-514-4 (1-103) x US-10-472-587-3 (1-312)

QY 1 MetThGluGlnAsnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr 20
DB 1 ATGACCGAGCAGACGAGCTGCAGCGGCTGCGCGGAGCTCGACGCTCGACGCGGACG 60
QY 21 LeuLeuAspThrValArgArgGlyLeuAspLeuGlyValArgIleAlaArgTyrIysSer 40
DB 61 CTCTGGACACGAGTGGCGCCCGCATCGACCTGGGTGCCGATCGGCGGTACAAAGTCC 120
QY 41 ArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAlaAla 60
DB 121 CGGACGCGGCTCCCGATGATGACGCGCGGCGGCTGATGCTCAAGACGAGGCGCC 180
QY 61 ArgTyrAlaIleAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspValIle 80
DB 181 CGCTACGCGCCGACGACGCGCTCGACGATCGTTCCTGTGAACCTCTACGACGCTGATC 240
QY 81 IleThrGluMetCysArgValGluAspLeuValMetSerArgGlySerLeuThrAlaGlu 100
DB 241 ATACCGAGATGTGCGCGCTGAGGACCTGTGATGACCGGAGAGCCTTGACGCGCGAG 300
QY 101 AspArgArg 103
DB 301 GACCGCGCG 309

RESULT 2
US-10-126-927-68
/ Sequence 68, Application US/10126927
/ Publication No. US20030082575A1
/ GENERAL INFORMATION:
/ APPLICANT: The Scripps Research Institute
/ APPLICANT: Schultz, Peter G
/ APPLICANT: Wang, Lei
/ APPLICANT: Anderson, John C
/ APPLICANT: Chin, Jason
/ APPLICANT: Liu, David R
/ APPLICANT: Magliery, Thomas
/ APPLICANT: Meggers, Eric L
/ APPLICANT: Mehl, Ryan A
/ APPLICANT: Pasternak, Miro
/ APPLICANT: Santoro, Stephen W
/ APPLICANT: Zhang, Zhiwen
/ TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
/ FILE REFERENCE: 54-000120US
/ CURRENT APPLICATION NUMBER: US/10/126,927
/ CURRENT FILING DATE: 2002-04-19
/ PRIOR APPLICATION NUMBER: US 60/285,030
/ PRIOR FILING DATE: 2001-04-19
/ PRIOR APPLICATION NUMBER: US 60/355,514
/ PRIOR FILING DATE: 2002-02-06
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: PatentIn version 3.1
```

```
/ SEQ ID NO 68
/ LENGTH: 3305
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-10-126-927-68

Alignment Scores:
Pred. No.: 5,456-63 Length: 3305
Score: 486.50 Matches: 103
Percent Similarity: 98.10% Conservative: 0
Best Local Similarity: 98.10% Mismatches: 0
Query Match: 94.65% Indels: 2
DB: 14 Gaps: 1

US-10-089-514-4 (1-103) x US-10-126-927-68 (1-3305)

QY 1 MetThGluGlnAsnGluLeuGln---ArgLeuArgAlaGluLeuAspAlaLeuAspG 19
DB 2990 ATGACCGAGCAGAACGAGCTGCAGGTTGCGGCTGCGCGGAGCTCGACGCTCGACG 3049
QY 19 YhrLeuLeuAspThrValArgArgGlyLeuAspLeuGlyValArgIleAlaArgTyrI 39
DB 3050 GACGCTTCTGACACGCTGCGCGCGCATGACCTCGGTGTCGCGATGCGCGGTACA 3109
QY 39 SerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValIysAspArgAl 59
DB 3110 GTCCCGGACGCGCTCCGATGATGACGCCCGCGGTCACTGCTCAAGACGAGGC 3169
QY 59 AlaArgTyrAlaIleAspHisGlyLeuAspGlySerPheLeuValAsnLeuTyrAspVa 79
DB 3170 CGCCCGCTACGCGCGCGGACGACGCGCTCGACGAATGTTCTGTGTAACCTCTACGAC 3229
QY 79 IleIleThrGluMetCysArgValGluAspLeuValMetSerArgGlySerLeuThrAl 99
DB 3230 GATCATCACGAGATGTGCGCGTGCAGGACCTGTGATGAGCCGAGAGCCTTGACGCG 3289
QY 99 aGluAspArgArg 103
DB 3290 CGAGGACGCGCGG 3302

RESULT 3
US-10-126-927-67
/ Sequence 67, Application US/10126927
/ Publication No. US20030082575A1
/ GENERAL INFORMATION:
/ APPLICANT: The Scripps Research Institute
/ APPLICANT: Schultz, Peter G
/ APPLICANT: Wang, Lei
/ APPLICANT: Anderson, John C
/ APPLICANT: Chin, Jason
/ APPLICANT: Liu, David R
/ APPLICANT: Magliery, Thomas
/ APPLICANT: Meggers, Eric L
/ APPLICANT: Mehl, Ryan A
/ APPLICANT: Pasternak, Miro
/ APPLICANT: Santoro, Stephen W
/ APPLICANT: Zhang, Zhiwen
/ TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
/ FILE REFERENCE: 54-000120US
/ CURRENT APPLICATION NUMBER: US/10/126,927
/ CURRENT FILING DATE: 2002-04-19
/ PRIOR APPLICATION NUMBER: US 60/285,030
/ PRIOR FILING DATE: 2001-04-19
/ PRIOR APPLICATION NUMBER: US 60/355,514
/ PRIOR FILING DATE: 2002-02-06
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 67
/ LENGTH: 12391
/ TYPE: DNA
/ ORGANISM: Plasmid pSci101, Streptomyces venezuelae papABC
US-10-126-927-67
```



```

US-10-433-556-20
: Sequence 20, Application US/10433556
: Publication No. US20040117872A1
:
: GENERAL INFORMATION:
: APPLICANT: Matrigne, Michel
: APPLICANT: Rippert, Pascal
: TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
: TITLE OF INVENTION: Said Herbicides
: FILE REFERENCE: 5500*120
: CURRENT APPLICATION NUMBER: US/10/433,556
: CURRENT FILING DATE: 2003-06-04
: NUMBER OF SEQ ID NOS: 30
: SOFTWARE: Patemlin version 3.2
: SEQ ID NO 20
:
: LENGTH: 1152
: TYPE: DNA
:
: ORGANISM: Ewinia herbiicola
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1119)
:
: US-10-433-556-20

```

Alignment Scores:	
Pred. No.:	7.11e-05
Score:	102.00
Percent Similarity:	50.00%
Best local Similarity:	27.55%
Query Match:	19.84%
DB:	19
Length:	1152
Matches:	27
Conservative:	22
Mismatches:	33
Indels:	16
Gaps:	2

US-10-089-514-4 (1-103) X US-10-433-556-20 (1-1152)

[illegible]

RESULT 7
 US-10-194-163-361/c
 : Sequence 361, Application US/10194163
 : Publication No. US20020172976A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Rose, Bruce Carter
 : TITLE OF INVENTION: PORPHYROMONAS GINGIVALIS POLYNUCLEOTIDES
 : AND USES THEREOF
 :
 : NUMBER OF SEQUENCES: 1120
 : CORRESPONDENCE ADDRESS:
 : ADDRESSER: MORRISON & FOERSTER
 : STREET: 755 PAGE MILL ROAD
 : CITY: PALO ALTO
 : STATE: CA
 : COUNTRY: USA
 : ZIP: 94304-1018
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Diskette
 : COMPUTER: IBM Compatible
 : OPERATING SYSTEM: Windows
 : SOFTWARE: FastSeq for Windows Version 2.0b

```

1      CURRENT APPLICATION DATA
2      APPLICATION NUMBER: US/10/194,163
3      FILING DATE: 04-Nov-2002
4      CLASSIFICATION: <Unknown>
5      ATTORNEY/AGENT INFORMATION:
6      NAME: Basu, Shantanu
7      REGISTRATION NUMBER: 43,318
8      REFERENCE/DOCKET NUMBER: 529282000101
9      TELECOMMUNICATION INFORMATION:
10     TELEPHONE: 650-813-5995
11     TELEFAX: 650-494-0792
12     TELEX: 706141
13
14     INFORMATION FOR SEQ ID NO: 361
15     SEQUENCE CHARACTERISTICS:
16         LENGTH: 850 base pairs
17         TYPE: nucleic acid
18         TOPOLOGY: double
19         STRANDEDNESS: double
20         MOLECULE TYPE: DNA (genomic)
21         HYPOTHETICAL: NO
22         ANTI-SENSE: UNKNOWN
23     ORIGINAL SOURCE:
24     ORGANISM: PORPHYROMONAS GINGIVALIS
25     FEATURE:
26     NAME/KEY: misc feature
27     LOCATION: 1..850
28     SEQUENCE DESCRIPTION: SEQ ID NO: 361
29     US-10-194-163-361

```

SEQUENCE DESCRIPTION:	SEQ ID NO:	361
US-10-194-163-361		

Alignment Scores:

Pred. No.:	0.00019	length:	850
Score:	98.00	Matches:	21
Percent Similarity:	56.34%	Conservative:	19
Best Local Similarity:	29.58%	Mismatches:	31
Query Match:	19.07%	Indels:	0
DB:	13	Gaps:	0

US-10-089-514-4 (1-103) x US-10-194-163-361 (1-850)

```

QY      |  GluGlnAsnGluIleuGlnIlnArgLeuArgAlaGluAspAlaLeuAspGlyThrLeuIleu  22
Db      |  214 CAAAGACGAAGAGCTGTATCTCTCTTGGCGCATCGAGATTGATCGAGATGAGAGATATGTG  155
QY      |  23  AspThrValArgArgArgIleAspLeuGluGlyValArgIleAlaArgTyrLysSerArgHis  42
Db      |  154 GAATTCCTGATGCTGGCGGATGCAAGTGGCATTCGAGATTAAGTTTGTTCAAAAAGAGAC  95
QY      |  43  GlyValProMetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyr  62
Db      |  94  AATCTGAGCTGTGGTTCAGAAATCTCCGCTACGAAACAATACAGCGCAACCTGTCCCTACT  35
QY      |
Db      |  63  AlaAlaAspHisGlyLeuAspGluSerPheLeu  73
Db      |  34  GCAGCCCTCTTAGGTGGACCAAAACATTATA  2

```

```

RESULT 8
US-10-381-779-40/C
: Sequence 40: Application US/10381779
: Publication No. US20030219798A1
: GENERAL INFORMATION:
: APPLICANT: Gokarn, Ravi R
: APPLICANT: Jessen, Holly
: APPLICANT: Zidwick, Mary Jo
: TITLE OF INVENTION: Isoprenoid Production
: FILE REFERENCE: 12904/002US1
: CURRENT APPLICATION NUMBER: US/10/381,779
: PRIOR APPLICATION NUMBER: 2003-03-28
: PRIOR FILING DATE: 2001-09-28
: PRIOR APPLICATION NUMBER: PCT/US01/30328
: PRIOR FILING DATE: 2000-09-29
: NUMBER OF SEQ ID NOS: 190
: SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 40
; LENGTH: 1833
; TYPE: DNA
; ORGANISM: Sphingomonas trueperi
US-10-381-779-40

```

Alignment Scores:

```

Pred. No.: 0.000943 Length: 1833
Score: 96.50 Matches: 27
Percent Similarity: 43.02% Conservative: 10
Best Local Similarity: 31.40% Mismatches: 48
Query Match: 18.77% Indels: 1
DB: 17 Gaps: 1

```

US-10-089-514-4 (1-103) x US-10-381-779-40 (1-1833)

```

QY 1 MetThrGlnGlnaAnglueGlnArgLeuAArgAlaGluLeuAapAlaLeuAapGlyThr 20
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 522 ATGACCGACGCGACG--CTGACCCCGCTATCGCCAGACATCGACATATCGATGCGCGC 466
QY 21 LeuLeuAapThrValaArgaArgGlyLeuApleuGlyValaArgIleAlaArgTyrLysSer 40
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 465 CTGCTCTTCATGCTCGCCGACGCTTCACAGTACCCAGCGCGGTGGCGCTACAGGCC 406
QY 41 ArgHisGlyValaProMetMetGlnProGlyArgValSerLeuValLysAapArgAlaAla 60
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 405 GAATCCGGCGTCCGCCGCCGATCCGCCGCGAGAGCGCAGATCGCCCGGCTGCCG 346
QY 61 ArgTyrAlaAlaAapHisGlyLeuAapGluSerPheLeuValaAsnLeuTyrAapValIle 80
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 345 GCGCTGGCGCAGGATGCCGATCTCGACCCGAAATTTCCGAGAAATTCCTGCGCTTCATC 286
QY 81 IleThrGluMetCysArg 86
    |||||:::|||||:::|||||:::|||||:::|||||:::
DB 285 ATCAGCAGGATGATCCGC 268

```

RESULT 9

```

US-10-156-761-6187
; Sequence 6187, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6187
; LENGTH: 387
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(387)
US-10-156-761-6187

```

Alignment Scores:

```

Pred. No.: 0.000742 Length: 387
Score: 91.00 Matches: 24
Percent Similarity: 45.68% Conservative: 13
Best Local Similarity: 29.63% Mismatches: 44
Query Match: 17.70% Indels: 0
DB: 15 Gaps: 0

```

US-10-089-514-4 (1-103) x US-10-156-761-6187 (1-387)

```

QY 6 GluLeuGlnArgLeuAArgAlaGluLeuAapAlaLeuAapGlyThrLeuLeuAapThrVal 25
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 58 GAATCGCCCGGCTGCGCCAGACATCGACATCGACCGCGCTCGCATCGATCTC 117
QY 26 ArgArgArgIleAapLeuGlyValaArgIleAlaArgTyrLysSerArgHisGlyValPro 45
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 118 GCCGAACGCTTCAAGTGCACCCAGCAGTCCGCCACCTCAAGCCGTACACGAGTGGC 177
QY 46 MetMetGlnProGlyArgValaArgValSerLeuValLysAapArgAlaAlaArgTyrAlaAasp 65
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 178 CCGCGCGACCGCGCGCGCCGACGACCGACATGCGCCGACGCGCGCTCGCGCGAGAC 237
QY 66 HisGlyLeuAapGluSerPheLeuValaAsnLeuTyrAapValIleIleThrGluMetCys 85
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 238 GCCAACTCGACCGCGCGCTTCGCGGAGAGCTGCTGCGCTTCATCTCCCGAGTTCATC 297
QY 86 Arg 86
    |||||:::|||||:::|||||:::|||||:::|||||:::
DB 298 CGC 300

```

RESULT 10

```

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

```

Alignment Scores:

```

Pred. No.: 955 Length: 9025608
Score: 91.00 Matches: 24
Percent Similarity: 45.68% Conservative: 13
Best Local Similarity: 29.63% Mismatches: 44
Query Match: 17.70% Indels: 0
DB: 15 Gaps: 0

```

US-10-089-514-4 (1-103) x US-10-156-761-1 (1-9025608)

```

QY 6 GluLeuGlnArgLeuAArgAlaGluLeuAapAlaLeuAapGlyThrLeuLeuAapThrVal 25
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 7459246 GAATCGCCCGGCTGCGCCAGACATCGACATCGACCGCGCTCGCATCGATCTC 7459305
QY 26 ArgArgArgIleAapLeuGlyValaArgIleAlaArgTyrLysSerArgHisGlyValPro 45
    |||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
DB 7459306 GCCGAACGCTTCAAGTGCACCCAGCAGTCCGCCACCTCAAGCCGTACACGAGTGGC 7459365
QY 46 MetMetGlnProGlyArgValaArgValSerLeuValLysAapArgAlaAlaArgTyrAlaAasp 65

```

```

Db      7459366 CCCGCCGACCCGCCGCCGAGCCAGATCGCCGCTGCCGCTCGCCGAGAAC 7459425
Qy      66 HieGlYleuAspJuserPheLeuValAsnLeuTyrAspValIleIleThrhGluMetCys 85
Db      7459426 GCCAACTCGACCCGCCGCTTCGCCGAGAACTGCTGGGCTTCATCATCGCCGAGTCAATC 7459485
Qy      86 Arg 86
Db      7459486 CGC 7459488

RESULT 11
US-10-137-310-3
; Sequence 3, Application US/10137310
; Publication No. US20030176675A1
; GENERAL INFORMATION:
; APPLICANT: Valentin, Henry E.
; APPLICANT: Malsky, Timothy A.
; TITLE OF INVENTION: Tyra Genes and Uses Thereof
; FILE REFERENCE: 16515.147
; CURRENT APPLICATION NUMBER: US/10/137,310
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 60/289,527
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-137-310-3

Alignment Scores:
Pred. No.:      0.00467      Length:      1122
Score:          90.00      Matches:      28
Percent Similarity: 49.49%      Conservative: 21
Best Local Similarity: 28.28%      Mismatches: 32
Query Match:    17.51%      Indels:    18
DB:             16      Gaps:          4

US-10-089-514-4 (1-103) x US-10-137-310-3 (1-1122)

Qy      6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
Db      10 GAATTGACCCGATACCGCATCAAAATTGATGAAGTGAACCGCTGGAATTATTATTA 69
Qy      26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
Db      70 GCGAAGCGTCTGGAACCTGTTGCTGAAGTGGCGGAGGTGAAAGCCGCTTGGACTGCCT 129
Qy      46 MetMetGlnProGlyArg--ValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64
Db      130 ATTATGTTCCGAGAGCGAGCATCTATGTTGGCCCTCGCTGTCAGAG--GCGGAA 186
Qy      65 AspHisGlyLeuAspJuserPheLeuValAsnLeuTyrAspValIleIleThrhGluMet 84
Db      187 GCTCTGGGTATACCGCAGATCTG----- 210
Qy      85 CysArgValGluSerLeuVal-----MetSerArgGluSerLeuThraGluAsp 101
Db      211 -----ATTGAGAGATGTTTGGCTGGGTGATGCGTGAATCTTACTCCAGTGAAGAAC 261

RESULT 12
US-10-433-556-18
; Sequence 18, Application US/10433556
; Publication No. US20040117872A1
; GENERAL INFORMATION:
; APPLICANT: Matringe, Michel
; APPLICANT: Rappert, Pascal
; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
; TITLE OF INVENTION: Said Herbicides
; FILE REFERENCE: 5500*120
; CURRENT APPLICATION NUMBER: US/10/433,556
```

```

; CURRENT FILING DATE: 2003-06-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1122)
US-10-433-556-18

Alignment Scores:
Pred. No.:      0.00467      Length:      1122
Score:          90.00      Matches:      28
Percent Similarity: 49.49%      Conservative: 21
Best Local Similarity: 28.28%      Mismatches: 32
Query Match:    17.51%      Indels:    18
DB:             19      Gaps:          4

US-10-089-514-4 (1-103) x US-10-433-556-18 (1-1122)

Qy      6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
Db      10 GAATTGACCCGATACCGCATCAAAATTGATGAAGTGAACCGCTGGAATTATTATTA 69
Qy      26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArgHisGlyValPro 45
Db      70 GCGAAGCGTCTGGAACCTGTTGCTGAAGTGGCGGAGGTGAAAGCCGCTTGGACTGCCT 129
Qy      46 MetMetGlnProGlyArg--ValSerLeuValLysAspArgAlaAlaArgTyrAlaAla 64
Db      130 ATTATGTTCCGAGAGCGGAGCATCTATGTTGGCCCTCGCTGTCAGAG--GCGGAA 186
Qy      65 AspHisGlyLeuAspJuserPheLeuValAsnLeuTyrAspValIleIleThrhGluMet 84
Db      187 GCTCTGGGTATACCGCAGATCTG----- 210
Qy      85 CysArgValGluSerLeuVal-----MetSerArgGluSerLeuThraGluAsp 101
Db      211 -----ATTGAGAGATGTTTGGCTGGGTGATGCGTGAATCTTACTCCAGTGAAGAAC 261

RESULT 13
US-10-380-132-7
; Sequence 7, Application US/10380132
; Publication No. US20030182679A1
; GENERAL INFORMATION:
; APPLICANT: SunGene GmbH & Co. KGaA
; TITLE OF INVENTION: Improved processes for vitamin E biosynthesis
; FILE REFERENCE: NAB445/2000
; CURRENT APPLICATION NUMBER: US/10/380,132
; CURRENT FILING DATE: 2003-03-11
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1238
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: gene
; LOCATION: (7)..(1232)
; OTHER INFORMATION: Cyra gene coding for bifunctional chorismate
; OTHER INFORMATION: mutase / prephenate dehydrogenase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(1143)
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)..(6)
; OTHER INFORMATION: restriction site linker
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1233)..(1238)
```



```
,
, LOCATION: (29298)..(29298)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36543)..(36543)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36551)..(36551)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (36636)..(36636)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (40808)..(40810)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44416)..(44416)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44905)..(44905)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (44975)..(44975)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (45593)..(45593)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (45732)..(45732)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (47036)..(47036)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51334)..(51334)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51602)..(51602)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51786)..(51786)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (51805)..(51805)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (55369)..(55369)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (65309)..(65309)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (65313)..(65313)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (80024)..(80024)
,
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (10091)..(10091)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (102696)..(102696)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (105121)..(105121)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (107248)..(107248)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (117136)..(117136)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (119750)..(119750)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (119924)..(119924)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (120038)..(120038)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (121344)..(121344)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (122167)..(122167)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (122316)..(122316)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (131340)..(131340)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (131360)..(131360)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (139910)..(139910)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (140398)..(140398)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (142750)..(142750)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (145058)..(145058)
, OTHER INFORMATION: n equals a, t, g or c
, FEATURE:
, NAME/KEY: misc_feature
, LOCATION: (145171)..(145171)
, OTHER INFORMATION: n equals a, t, g or c
```

```

FEATURE:
NAME/KEY: misc feature
LOCATION: (145942) . (145942)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (147197) . (147197)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841) . (150841)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500) . (152500)
OTHER_INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc_feature
LOCATION: (152530) . (152530)

```

Pred. No.:	27	Length:	1830122
Score:	88.50	Matches:	17
Percent Similarity:	67.92%	Conservative:	19
Best Local Similarity:	32.08%	Mismatches:	16
Query Match:	17.22%	Indels:	1
DB:	17	Gaps:	1

QY	Db	QY	Db	QY	Db
7	1370466	27	1370526	47	1370586
LeuGlnArgLeuArgIleGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArg	CTTAAAGATTACGTTCTGTAATGATTGCGCTTATGCGCAACTTATCCACTTTTGCT	ArgArgIleLeuLeuGluValArgIleAlaArgIleArgSerArgHisGluValProMet	AAACGCTTACGTTGCTTTCTCAAGTCGCGTAAAGTAAACATCAACATGATTTACTATT	MetGlnProGlyArg--ValSerLeuValIleAspArg	TATGCGCCAGAACGTGAATATGCAATCTCCAAACAGCT
26	1370525	46	1370585	58	1370624

Search completed: October 6, 2005, 01:50:33
Job time : 5167.67 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 1126.79 seconds
(without alignments)
3479.464 Million cell updates/sec

Title: US-10-089-514-4

Perfect score: 514
Sequence: 1 MTEQNELQRLRAELDALDGT.....MCRVEDLVMSRESLTAEADR 103

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO.spool.h/US10089514/runat_04102005_105744_8043/app_query.fasta_1.1621
-DB=EST -QFMT=fastap -SUFFIX=rc -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0
-UNITS-bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS-human40.cdi -LIST=45
-DOCALLGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=prco -NOR=ext -HEA=SIZE=500 -MINLEN=2000000000
-USRR=US10089514_@CGN_1_1_6050_@runat_04102005_105744_8043 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
EST: *
1: gb_est1.*
2: gb_est2.*
3: gb_hic.*
4: gb_est3.*
5: gb_est4.*
6: gb_est5.*
7: gb_est6.*
8: gb_gss1.*
9: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	85.5	16.6	806	9	CL688507 PRI0149C
2	82	16.0	846	7	CN822873 Oa.sp1bn
3	79.5	15.5	600	8	BZ895453 NaEP11.01
4	79	15.4	471	2	AM637288 b156h07.w
5	78	15.2	431	9	CL674789 PRI0113a
6	77.5	15.1	509	4	BM326074 PIC1_65_E
7	77	15.0	692	8	BH454175 BOCI86TF
8	76.5	14.9	913	5	BU915943 AGENCOURT
9	75.5	14.7	627	7	CN894351 010625AAV

C	10	75	14.6	529	4	BM595564	BM595564 170006874
C	11	75	14.6	644	8	BH654655	BH654655 BOHTE10TF
C	12	75	14.6	756	8	BH248862	BH248862 BOGAM09TF
C	13	74	14.4	312	6	CD602770	CD602770 RK354A2F1
C	14	74	14.4	325	6	CD602824	CD602824 RK354A4G1
C	15	74	14.4	535	7	CF857652	CF857652 pBNY0021F
C	16	74	14.4	536	2	BE583221	2-3F-MY P
C	17	74	14.4	559	5	BX601867	BX601867 BX601867
C	18	74	14.4	615	9	CL708793	CL708793 OR_BBA003
C	19	74	14.4	739	1	AL860917	AL860917 w156B07.x
C	20	74	14.4	775	8	BZ577997	BZ577997 man2_5671
C	21	74	14.4	785	9	CL727759	OR_BBA005
C	22	73.5	14.3	368	6	CD061523	CD061523 MA1-0020P
C	23	73.5	14.3	569	1	AA941215	AA941215 LBD5234.5
C	24	73.5	14.3	595	7	CF429729	CF429729 PH1_24_H0
C	25	73.5	14.3	624	7	CV233416	CV233416 WS01210.B
C	26	73.5	14.3	633	7	CV238063	CV238063 WS0125.B2
C	27	73.5	14.3	662	7	CV237417	CV237417 WS01227.B
C	28	73.5	14.3	685	7	CV237350	CV237350 WS01227.B
C	29	73.5	14.3	691	7	CV236915	CV236915 WS01225.B
C	30	73.5	14.3	727	9	CG840622	Ybhw1235
C	31	73.5	14.3	909	4	BG343273	BG343273 HYSME5000
C	32	73	14.2	657	8	BH934632	BH934632 odt87C04.
C	33	73	14.2	693	8	BH576490	BH576490 BOGGA03TF
C	34	73	14.2	769	8	BH250514	BH250514 BOGGAU29TF
C	35	73	14.2	789	8	BH436424	BH436424 BOHJ044TF
C	36	73	14.2	908	7	CF700884	CF700884 CCAF289TO
C	37	73	14.2	1019	7	CF706740	CF706740 CCA240TO
C	38	72.5	14.1	460	7	CO312074	CO312074 EK227832.
C	39	72.5	14.1	505	1	AA804082	AA804082 GM12763.5
C	40	72.5	14.1	530	2	BF494794	BF494794 AT03534.5
C	41	72.5	14.1	562	1	AL113596	AL113596 GH09960.5
C	42	72.5	14.1	572	4	B1355453	B1355453 CM30880.5
C	43	72.5	14.1	674	1	AI106885	AI106885 GH06135.5
C	44	72.5	14.1	712	2	BF493712	BF493712 AT02044.5
C	45	72.5	14.1	738	9	CNS06UHZ	AL415869 T3 end of

ALIGNMENTS

RESULT 1
CL688507
LOCUS
DEFINITION
PRI0149C_P10_2 - PRI0149C_BR (806) Mixed stage fosmid library of P. pacificus var. California *Pristionchus pacificus* genomic, genomic survey sequence.

ACCESSION
CL688507
VERSION
CL688507.1 GI:50197585

SOURCE
GSS.
ORGANISM
Pristionchus pacificus

REFERENCE
AUTHORS
Srinivasan J., Otto G.W., Kahlow U., Geisler R. and Sommer R.J.

TITLE
ApprAB: an AcetB database for the nematode satellite organism

JOURNAL
Pristionchus *pacificus*
Nucleic Acids Res. 32 (1), D421-D422 (2004)

COMMENT
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498

EMAIL: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.

Seq primer: T7
Class: fosmid end.

Location/Qualifiers
1..806
/organism="Pristionchus pacificus"
/mol_type="genomic DNA"

FEATURES

source

```

/serain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus
var. California"
/note="Vector: pBplfos-5 Fosmid vector"

```

Alignment Scores:	
Pred. No.:	0.608
Score:	85.50
Percent Similarity:	59.26%
Best Local Similarity:	35.19%
Query Match:	16.63%
DB:	9
Length:	806
Matches:	
Conservative:	13
Mismatches:	21
Indels:	
Gaps:	1

US-10-089-514-4 (1-103) x CL688507 (1-806

Oy 6 GluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrVal 25
 ||||| ||||| ::::: ||| ||||| :::
 Db 640 GAATTCAGCCGCATTACGGCATCAAAATGTATGATGATTAAGCGCGCTGTAATTATTTA 699

Qy 26 ArgArgArgIleAspLeuGlyValArgIleAlaArgTyrIlysserAghIsclValPro 45
 ::|||:::||||| ::| ||||| |||:::|||||
 Db 700 GCGAAGCGTGTGGAAGTGGTCTGAAGTGGGCGAGGTGAATAAGCCGCTTTGGACTGGCT 759

QY	46 MetMetGlnProGlyArg---ValSerLeuValLysAspArg	58
	::: :::::	
Db	760 ATTATGTTCCGAGCGGAGCAGCATCTATGTTGGCCTCGCGG	801

RESULT 2
CN822873/c
LOCUS
DEFINITION
CN822873
Oa_splbn_02E15_M13reverse Sheep spleen/brain pSpoc1 library Ovis
aries cDNA clone Oa_splbn_02E15 5' mRNA sequence.
accession

VERSION	CN822873.1	GI:47950942
KEYWORDS	EST.	
SOURCE	Ovis aries (sheep)	
ORGANISM	Ovis aries	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Caprinae; Ovis.				
1 (bases 1 to 846)	Gossner, A. and Hopkins, J.	Ovine spleen/brain cDNA library	Unpublished (2004)	
	Contact: J Hopkins			

Veterinary Biomedical Sciences
 University of Edinburgh
 Summerhall Square, Edinburgh, EH9 1QH
 Email: j.hopkins@ed.ac.uk
 Plate: 02 row: E column: 15
 Seq primer: M3reverse
 High quality sequence start: 4
 High quality sequence stop: 549.

```

FEATURES
source
location/Qualifiers
1..846
/organism="Ovis aries"
/mol_type="mRNA"
/db_xref="taxon:9940"
/clone="Oa_splbn_02815"
/clone_lib="Sheep spleen\brain p5port1 library"
/note="Vector: p5port1"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	1 94
Score:	82.00
Percent Similarity:	46.30%
Best Local Similarity:	29.63%
Query Match:	15.95%
DB:	7
Gaps:	
	846
Length:	
Matches:	32
Conservative:	18
Mismatches:	28
Indels:	30
Gaps:	7

US-10-089-514-4 (1-103) X CN822873 (1-846)

QY 4 GLIAsngIuLeugIArgLeuArgIaGluLeuAspAlaLeuAspGlyThreu----- 21
||| :: ||||| ::||| ||||| ::
Db 689 CAGCGCGATCAGACGCGCAACAGCGCCACCGCCACCA---GATGGGGGAATTCCACTC 633

Qy 22 -----LeuAspPfrValArgArgArgIleAspLeuGlyVal-----ArgIleAlaArg 37
||| :: |||::||| ||| |||::|||
Db 632 GTGCATCTTGGCGCTGGCGGACGGGATCATTCATGTAAGTGCAGCGGTGTCGGCAC 573

```

Dy      38 TyrLysSerArgHisGlyVal-----PrometMet----- 47
      :: ||| ||||| ||| :::::
Db      572 CATCTGGGCTGGCTTTGGCGCTTGGCCGACCCCACTGCTGTAGCGCGCGCCGACATCGGSA 5133

```

```

oy 48 -----GimProGIvArvalseIreUuallYsaspargalaaIaIrIyIaIaIa 64
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 512 GTAGCGGAACATCCAGGCGCGCTGCGGCTG-----CACGATGC 474

```

Dy 65 ASPTATGTCAGTGCCTGATCAGGCTC-----CTTTCCGGACCGTCGTCAAGAACTT 423

Db 473 GACCATGCCCTGATCAGGCTC-----CTTTCCGGACCGTCGTCAAGAACTT 423

cy cysatgvaagcuaabpuenvalmec 92
:::||| |||||:::
db 422 CATCTGCTCGAACAGCTTGTGGTA 399

RESULT 3	BZ895453	LOCUS	DEFINITION
BZ895453	600 bp	DNA	linear
NRP11_0120	Na pUC18	library	Natrialba asiatica genomic 5', genomic
survey	sequence.		

ACCESSION	B2895453	
VERSION	B2895453.1	GI:33345929
KEYWORDS	GSS.	
SOURCE	Natrialba asiatica	
ORGANISM	Natrialba asiatica	

Archaea; Euryarchaeota; Halobacteria; Halobacteriales;
Halobacteriaceae; Natrionalba.
1 (bases 1 to 500)

REFERENCE	1 (bases 1 to 600)
AUTHORS	Gao, Y., Roach, J., Glusman, G., Baliga, N. S., Deutsch, K., Pan, M.,
TITLE	Dassarma, S., Ng, W. V. and Hood, L.
JOURNAL	Low-pass Sequencing for Microbial Comparative Genomics
COMMENT	Unpublished (2003)
	Contact: Gao Y

Institute for Systems Biology
1441 North 34th Street, Seattle, WA 98103, USA
Tel: 206 732 1412
Fax: 206 732 1299
Email: ygc00@systemsbiology.org
Seg primer: M13 forward
Class: shotgun.

```

FEATURES
source
1..600
location/Qualifiers
/organism="Natrialba asiatica"
/mol_type="genomic DNA"
/strain="ATCC 700177"
/db_xref="taxon:64602"
/clone_lib="Na pUC18 library"
/note="Vector: pUC18. Site 1: SmaI; A shotgun library was
constructed from Natrialba asiatica genomic DNA using
pUC18/SmaI/BAP plasmid"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	2.78
Score:	79.50
Percent Similarity:	51.06%
Best Local Similarity:	23.40%
Query Match:	15.47%
DB:	8
Length:	600
Matches:	22
Conservative:	26
Mismatches:	41
Indels:	5
Gaps:	2

US-10-089-514-4 (1-103) X B2895453 (1-600)

Qy 3 GluGlnAnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeu 22
Db 28 GAGCAGATGATCTCGAACAACCTCCGAGAGATGATGACGATCATCGAGATGCTC 87
Qy 23 AspThrValArgArgAlaGlyLeuAspLeuGlyValArgIleAlaArgIlyrLySerArgHis 42
Db 88 GAACGTGATCCGCGCAGCACTACGTCCGAGATACGATTCGACGGTCAAGCAGCAGAG 147
Qy 43 G1ValPrometMetGlnProG1ArgValSerLeuValysAspArgAlaAlaArgTyr 62
Db 148 GACCTGCCGACGACCGCAGACCATGATGACAGTAAATGACAGCAGCGGCGCAGAAC 207
Qy 63 AlaAlaAspHisGlyLeuAspGluSerPheLeuValAsnLeuTyrAspValIleIleThr 82
Db 208 GCAGAGACGATTGCGATCGACGCGAATCTGTGTAAGCGCATCTTCGCTATTGATC--- 264
Qy 83 GluMetCysArgValGluAspLeuValMetSerArgGluSer 96
Db 265 GAACGTGAACAAGTTGAG-----CAGCGCAGTCC 294
RESULT 4
AM637288 471 bp mRNA linear EST 26-APR-2001
LOCUS b156h07.w1 Blackshear/Soares normalized Xenopus egg library Xenopus
DEFINITION laevis cDNA clone PBX0056H07 5', mRNA sequence.
ACCESSION AM637288
VERSION AM637288.1 GI:7394396
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
Xenopodidae; Xenopus; Xenopus.
1 (bases 1 to 471)
Blackshear,P.J., Lai,W.S., Thorn,J.M., Kennington,E.A., Steffa,N.G.
Jr., Moore,D.T., Bouffard,G.G., Beckstrom-Sternberg,S.M.,
Touchman,J.W., Bonaldo,M.F. and Soares,M.B.
The NIHES Xenopus maternal EST project: interim analysis of the
first 13,879 ESTs from unfertilized eggs
Gene 267 (1), 71-87 (2001)
21211403
11311557
JOURNAL Contact: Perry J. Blackshear
MEDLINE Office of Clinical Research and Laboratory of Signal Transduction
PUBMED National Institute of Environmental Health Sciences
COMMENT A2-05 NIHES, 101 Alexander Drive, Research Triangle Park, NC 27709,
USA
Tel: 919 541-4899
Fax: 919 541-4571
Email: black009@niehs.nih.gov
Clone is available through Research Genetics, Inc., 2130 Memorial
Parkway, Huntsville, AL 35901
phone 800-533-4363 ext.cdna, fax 256-536-9016 att:cdna, email
cdna@resgen.com
DNA Sequencing and analyses performed by National Institutes of
Health Intramural Sequencing Center (NISC).
PCR Primers
FORWARD: TGTAAACGACGCGCCAGT
BACKWARD: CAGGAACAGCTATGACC
Plate: 0056 row: H column: 07
Seq primer: T7 primer.
Location/Qualifiers
1..471
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="PBX0056H07"
/sex="female"
/tissue_type="unfertilized egg"
/cell_type="unfertilized egg"
/dev_stage="unfertilized egg"
/lab_host="DH108"
/clone_lib="Blackshear/Soares normalized Xenopus egg"

library"
/note="Vector: pT7T3-Pac; Site 1: EcoRI; Site 2: NotI;
polyA-selected mRNA was prepared from unfertilized Xenopus
laevis eggs. The library was constructed in the vector
pT7T3-Pac as described in Bonaldo, M.F., Lennon, G. and
Soares, M.B. 'Normalization and subtraction: two
approaches to facilitate gene discovery', Genome Research
6:791-806, 1996. The first strand synthesis used a
NotI-dt18 primer; double stranded cDNAs were ligated to
EcoRI adapters, digested with NotI, and directionally
cloned into the NotI and EcoRI-digested pT7T3-Pac vector.
The library contained approximately 7.2 x 10⁵
recombinants, with average insert sizes of 1-1.5 kb."

ORIGIN
Alignment Scores:
Pred. No.: 2.4 Length: 471
Score: 79.00 Matches: 28
Percent Similarity: 50.48 Conservative: 25
Best Local Similarity: 26.67 Mismatches: 38
Query Match: 15.37 Indels: 14
DB: 2 Gaps: 4
US-10-089-514-4 (1-103) x AM637288 (1-471)
Qy 4 GlnAnGluLeuGlnArgLeuArgAlaGluLeuAspAlaLeuAspGlyThr----- 20
Db 10 AGGCTGAAATTAATAAGCTGAAGGCCAGCTTCTCACAGCGTGAAGTCCACCAAGAC 69
Qy 21 LeuLeuAspThrValArgArgArgIle-----AspLeuGlyValArgIle 35
Db 70 GTCCTGATGGATGATGAGAAAGATCTCACAGCTCAGAGAGTTCACCCAGAGAT 129
Qy 36 AlaArgTyrLySerArgHisGlyValPrometMetGlnProG1ArgValSerLeuVal 55
Db 130 ATCAGATGCGCTTCCAGTCAGCAGCATGCTGAGAGAG-----AAATTAAGCCGAA 183
Qy 56 LysAspArgAlaAlaArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeuValAsn 75
Db 184 CGGACAGAGCGCAGAGAGAGAGGCTGCAAGCAGAGATCTCTTAATGCTCTGGAAAC 243
Qy 76 LeuTyrAspValIleIleThrGluMetCysArgValGluAspLeuValMetSerArgGlu 95
Db 244 TTGAGCAG-----GAGATTGCCGTTATCGGAATCTTTCAGACGCAAGAG 291
Qy 96 SerLeuThrAlaGlu 100
Db 292 CCGGATGCGCGCAG 306
RESULT 5
CL674789 431 bp DNA linear GSS 09-JUL-2004
LOCUS PRI0113a.E03.2 - PRI0113a.BR (431) Mixed stage fosmid library of P.
DEFINITION pacificus var. California Pristionchus pacificus genomic, genomic
survey sequence.
ACCESSION CL674789
VERSION CL674789.1 GI:50178309
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus;
1 (bases 1 to 431)
Srinivasan,J., Otto,G.W., Kahlow,U., Geisler,R. and Sommer,R.J.
ApbADB: an Acedb database for the nematode satellite organism
Pristionchus pacificus
Nucleic Acids Res. 32 (1), D421-D422 (2004)
Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de
This library was generated at Caltech, Pasadena, USA and end
sequenced at Vancouver, Canada.

FEATURES

Location/Qualifiers

```

/organism="Pristionchus pacificus"
/mol_type="genomic DNA"
/strain="California"
/db_xref="taxon:54126"
/clone_lib="Mixed stage fosmid library of P. pacificus"
var. California"
/note="Vector: pBspIos-5 Fosmid vector"

```

ORIGIN

Alignment Scores:	
Pred. No.:	2.95
Score:	78.00
Percent Similarity:	45.68%
Best Local Similarity:	28.40%
Query Match:	15.18%
DB:	9
length:	433
Matches:	23
Conservative:	14
Mismatches:	14
Indels:	0
Gaps:	0

US-10-089-514-4 (1-103) X CL674789 (1-431)

QY 1 MetThrGlnGlnAsnGlnIleuGlnIleArgLeuGlnGlnIleAspAlaIleuAspGlyThr 20
 143 ATACACTCGGAACCCGCTTACTGGCGCTCGGACGAGAAATACAGCGCGCTGATGAAAA 202
 QY 21 LeuLeuAspThrValArgArgArgIleAspLeuGlnValArgIleAlaArgTyrLeuSer 40
 203 TTATTTGACATTACTCGACAGCGGCGGCGCAACTGCGCGCTGAGTGGGAAAGCCAAACTG 262
 Db 41 ArgHisGlyValProMetMetGlnProGluArgValSerLeuValIleAspArgAla 60
 263 CTCTCGCATCGCCCGGTACGTGATTTTGTGTGACCCGATTTACTGGAAAGATTATTT 322
 Db 61 ArgTyrAlaAlaAspHisGlyLeuAspArgIleSerPheLeuValAsnLeuTyrAspValIle 80
 323 ACGCTCGGTAAAGCGACCATCTGGACGGCCCATTCATTACTCGCGCTGTTCCAGCTATC 382
 QY 81 Ile 81
 Db 383 ATT 385

RESULT 6	LOCUS	DEFINITION
BM326074/c	509 bp	mRNA linear
BM326074	509 bp	mRNA linear
PI1C1 65 E01.b1 A002	Pathogen-infected compatible 1 (PI1C1)	Sorghum
bicolor cDNA, mRNA sequence.		

REFERENCE 1 (bases 1 to 509)
Cordonier-Pratt, M.-M., Gingle, A., Fang, G.C., Dean, R., Wing, R.,
Sudman, M. and Pratt, L.H.

JOURNAL Unpublished (2002)
COMMENT Contact: Cordonnier-Pratt MM

Sequences have been trimmed to exclude PolyA, vector, and regions

below three quality 16. The threshold for highest quality sequence is 20. Three-prime sequences, which are obtained with POLYTMX or T7 sequencing primer, are presented as the reverse complement. Seq primer: JEN REV
High quality sequence stop: 433
POLYA=No.

FEATURES	Location/Qualifiers
source	1. .509

```

/organism="Sorghum bicolor"
/mol_type="rRNA"
/cultivar="BTx623"
/db_xref="taxon:4558"
/tissue_type="Leaves"
/dev_stage="4-week-old seedlings infected with
Colletotrichum graminitcola"
/clone_1b="Pathogen-infected compatible 1 (P1c1)"
/notes="Vector: pBluescript II SK(-) from Lambda Zap II;
Site 1: XhoI; Site 2: EcoRI; Four-week-old sorghum
seedlings were sprayed with spore suspension prepared from
3-week-old PM421, a sorghum isolate of the antherose
pathogen Colletotrichum graminitcola. Inoculated plants
were kept in a 25 C dark growth chamber with 100% relative
humidity for 24 hr, followed by 12/12 hr of light/dark
cycle at 25 C with 90% relative humidity for another 24
hr. All leaves were harvested and quick frozen with liquid
nitrogen and stored in a -80 C freezer. The library was
made from poly-A RNA in the cloning vector lambda Zap II.
Clones to be sequenced were prepared by mass excision.
WARNING: While most or all ESTs are expected to derive
from the host plant, no effort was made to eliminate ESTs
deriving from the pathogen."

```

ORIGIN

Alignment Scores:	
Pred. No.:	4, 24
Score:	77.50
Percent Similarity:	44.04%
Best Local Similarity:	28.44%
Query Match:	15.08%
DB:	4
	509
Length:	Matches:
	31
	Conservative:
	17
	Mismatches:
	34
	Indels:
	27
	Gaps:
	5

US-10-089-514-4 (1-103) X BM326074 (1-509)

Qy		4	GlnAsnGluLeuGlnArg-----LeuArgAlaGluLeuAspAlaLeuAsp-----	18
Db		361	CAAAAGCAAGCTCTCCGGCGAAGACCTCCGACAGCACCGCGGTGGAGTGCATGATGTC	302
Qy		19	-----GlyThrLeuLeuAspThrValArgArgArgIleAspLeuGlyValArgIleAla	36
Db		301	TTCCTCGGGCGGCTTGACGAAGCTCTTCCAGGGAGACTGCACGCGCGTTTCAC-----	248
Qy		37	ArgTyLysSerArgHisGlyValProMetMetGlnProGlyArgValSerLeuValLys	56
Db		247	-----GGCAGAAGCGGACCATTCGAC	227
Qy		57	AspArgAlaAlaArgTYrAlaAlaAspHis-----GlyLeuAspGluSerPheLeuVal	74
Db		226	GGAACCCGCCGACTGACGAGCGCGGAGCAGCGCTCGAGTCGATGTCATTCTTCACG	167
Qy		75	AsnLeuTYrAspValIleIleThrGluMetCysArgValGluAspLeuValMetSerArg	94
Db		166	CGGCTCGAAGAGTCTCTTCACG-----CGACTCGACGACCGCGCTTCACGGCCGA	116
Qy		95	GluSerLeuThrAlaGluAspArgArg	103
Db		115	ACGGGACCATTGACGAGAACCGCCGA	89

RESULT 7			
BH454175/c			
LOCUS	BH454175	692 bp	DNA
DEFINITION	BOG1X86TF BOG1	Brassica oleracea	genomic clone BOG1X86, genomic survey sequence.
ACCESSION	BH454175		

VERSION BH454175.1 GI:17639886
 KEYWORDS GSS.
 SOURCE Brassica oleracea
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
 REFERENCE 1 (bases 1 to 692)
 AUTHORS Town, C.D., Van Aken, S., Utterback, T., Koo, H. and Fraser, C.M.
 TITLE Whole genome shotgun sequencing of Brassica oleracea
 JOURNAL Unpublished (2001)
 COMMENT Other GSSs: BOGIX86TR
 Contact: Chris Town
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA.
 Tel: 301-838-3523
 Fax: 301-838-0208
 Email: cdtown@tigr.org
 DNA is from a doubled haploid provided by Tom Osborn.
 Seq primer: TP
 Class: sheared ends.
 FEATURES
 source
 Location/Qualifiers
 1..692
 /organism="Brassica oleracea"
 /mol_type="genomic DNA"
 /strain="TO1000DH3"
 /db_xref="taxon:3712"
 /clone="BOGIX86"
 /clone_1lb="BOGI"
 /note="Vector: PHOS1, Site 1: BstXI; 2-3 kb sheared genomic DNA inserted into PHOS1 using BstXI linkers"
 ORIGIN
 Alignment Scores:
 Pred. No.: 7.27 Length: 692
 Score: 77.00 Matches: 33
 Percent Similarity: 42.59% Conservative: 13
 Best Local Similarity: 30.56% Mismatches: 46
 Query Match: 14.98% Indels: 16
 DB: 8 Gaps: 4
 US-10-089-514-4 (1-103) x BH454175 (1-692)
 QY 6 GtLcUcLhArgleuAArglAglUleuAspAlaLeuAepgLythrVal 25
 Db 363 CAACCTGAACAGTGTGAGGCTGATCTGCAGAGGCTGCACGGGTGATGACCCGACGCT 304
 QY 26 ArgArgArglleAepLeuGlyValArglleAlaArgTyrLysSerArgHlsGlyValPro 45
 Db 303 GCTCGACGCGACAGATT-----AAAGCTCCGCTGATTCTCTCGACGACGACCGCTCCG 250
 QY 46 MetMetGlnProGlyArgValSerLeuValLysAspArgAlaAlaArgTyrAlaAlaAsp 65
 Db 249 GAGCTTGAACAGAAAAACACCGATCTCGAGAGTGAGAGGGCTTCA---CTCGCTGCCCGC 193
 QY 66 HisGlyLeuAepgLeuSerPheLeuValAsnLeuTyrAspValIlellethrgLumetCys 85
 Db 192 CAC-----GAGCGAGAGATGAACGTTCTGAGAACTCCAGAACTTGGAGGTGACG 142
 QY 86 ArgValGluAspLeuValMetSerArgGluSerLeuThrAla----- 99
 Db 141 AGGAAAGGGGGAGAGTAGAGCGGGAATGATGCCCAAGGCTAATCTTGTTCAGG 82
 QY 100 -----GluAspArgArg 103
 Db 81 ATTGGTTCTCGAGAGGATGCTCGG 58
 RESULT 8
 BU915943 913 bp mRNA linear EST 17-OCT-2002
 LOCUS AGENCOURT 10492829 NICHX XGC OOI Xenopus laevis cDNA clone
 DEFINITION IMAGE:6642729 5', mRNA sequence.
 ACCESSION BU915943

VERSION BU915943.1 GI:24097857
 KEYWORDS EST.
 SOURCE Xenopus laevis (African clawed frog)
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae; Xenopodinae; Xenopus; Xenopus.
 REFERENCE 1 (bases 1 to 913)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncigap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Martha Reibert, Steven L. Klein, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Agencourt Bioscience Corporation
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
 Plate: LLNL4228 row: j column: 19
 High quality sequence stop: 666.
 Location/Qualifiers
 1..913
 /organism="Xenopus laevis"
 /mol_type="mRNA"
 /db_xref="taxon:8355"
 /clone="IMAGE:6642739"
 /clone_1lb="DH10B (phage:resistant)"
 /note="Vector: pCMV-SF0R16, Site 1: NotI, Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.2 kb. Constructed by Life Technologies."
 ORIGIN
 Alignment Scores:
 Pred. No.: 12 Length: 913
 Score: 76.50 Matches: 25
 Percent Similarity: 50.00% Conservative: 25
 Best Local Similarity: 25.00% Mismatches: 39
 Query Match: 14.88% Indels: 11
 DB: 5 Gaps: 3
 US-10-089-514-4 (1-103) x BU915943 (1-913)
 QY 1 MetThGluGlnAsnGluLeuGlnArgLeuAArglAglUleuAspAlaLeuAepgLythr 20
 Db 508 TTGCTGAAATAATTAAGCTGAAGGCCAGCTCTCACAGCTGAGCTGACCCCAAGAGC 567
 QY 21 LeuLeuAspThrValArgArglle-----AspLeuGlyValArglle 35
 Db 568 GTGCTGATGGATGATGAGGAAGAGATCTCACAGCTCAGAGATTACACGAGAGAT 627
 QY 36 AlaArgTyrLysSerArgHlsGlyValProMetMetGlnProGlyArgValSerLeuVal 55
 Db 628 ATCAGATGGCTTCCGTCAGAGCATCACTGGGAGAG-----AAATTAAAGCCGA 681
 QY 56 LysAspArgAlaAlaArgTyrAlaAlaAspHlsGlyLeuAepgLeuSerPheLeuValAsn 75
 Db 682 CGGCAAGAGGCGAGAGAGGCTGCAACACAGAGATGCTTAAATGCTTGGAAAC 741
 QY 76 LeuTyrAspValIlellethrgLumetCysArgValGluAspLeuValMetSerArgGlu 95
 Db 742 TTGAG-----CCAGGAATTCGCCGTTATCGGAATCTCTTACAGACGAAG 789
 RESULT 9
 CN894351/c 627 bp mRNA linear EST 04-JUN-2004
 LOCUS CN894351 O10625AAVA002796HT (AAVA) Royal Gala 126 DAFB fruit cortex Malus x domestica cDNA clone AAVA002796, mRNA sequence.
 ACCESSION CN894351

VERSION	CN894351.1	GI:48280593
KEYWORDS	EST	
SOURCE	Malus x domestica (cultivated apple)	
ORGANISM	Malus x domestica	
REFERENCE	McIntyre, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B., McArthur, S., Newcomb, R., Ross, G., Snowden, K., Walton, E. and Yauk, Y.	
AUTHORS	McIntyre, L., Bowen, J., Crowhurst, R., Gleave, A., Janssen, B., McArthur, S., Newcomb, R., Ross, G., Snowden, K., Walton, E. and Yauk, Y.	
TITLE	HortResearch Apple EST Project	
JOURNAL	Unpublished (2004)	
COMMENT	Contact: Gleave, A. Sequencing Facility The Horticulture and Food Research Institute of New Zealand Ltd 120 Mt Albert Rd, Mt Albert, Auckland, New Zealand Tel: 00 64 09 815 4200 Fax: 00 64 09 815 4201 Email: est@hortresearch.co.nz.	
FEATURES	location/Qualifiers	
source	1..627	
	/organism="Malus x domestica"	
	/mol_type="mRNA"	
	/db_xref="taxon:3750"	
	/clone="AY002796"	
	/tissue_type="fruit cortex"	
	/dev_stage="126 days after full bloom"	
	/clone_11b="(AAY) Royal Gala 126 DAFB fruit cortex"	
	/note="Vector: pBluescript SK(-); Library sequenced by Genes Research & Development"	
ORIGIN		
Alignment Scores:		
Pred. No.:	10.3	Length: 627
Score:	75.50	Matches: 26
Percent Similarity:	58.33%	Conservative: 9
Best Local Similarity:	43.33%	Mismatches: 20
Query Match:	14.65%	Indels: 5
DB:	7	Gaps: 4
US-10-089-514-4 (1-103) x CN894351 (1-627)		
OY	11	ArgAlaGluLeuAspAlaLeuAspGlyThrLeuLeuAspThrValArgArgIleAsp 30
DB	266	AGAGAGGAGGCTCGAGATTCTT---GGGAGAGCTGTGGCTCCGCGAGGATCTTAC 210
OY	31	---LeuGlyValArgIleAlaArgTyrTyrSerArgHisGlyValProMetMetGlnPro 49
DB	209	GAGCTGGGTGCACGC-----CGGCGACGAAATCTCTCGGGGATACCGCTCGGCGACGCTT 156
OY	50	GlyArgValSerLeuValIleAspArgAlaAlaArg---TyrAlaAlaAspHisGlyIleu 68
DB	155	GACCAAGACAGGCTTCGACAGCGCGAGTGGCCGCGACGAGTACCGCGGAATTGTGCACTC 96
RESULT 10		
BM595564/c		
LOCUS	BM595564	529 bp mRNA linear EST 25-FEB-2002
DEFINITION	17000687475785 A.Gam.ad.cDNA, blood1 Anopheles gambiae cDNA clone	
ACCESSION	1960049718069 5', mRNA sequence.	
VERSION	BM595564	
KEYWORDS	BM595564.1 GI:18891426	
SOURCE	EST	
ORGANISM	Anopheles gambiae (African malaria mosquito)	
	Anopheles gambiae	
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Preyoptera;	
	Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;	
	Anopheles.	
REFERENCE	1 (bases 1 to 529)	
AUTHORS	Holt, R.A., Lin, J.-J., Murphy, S.D., Evans, C.A., Kraft, C.L.,	
	Charlbb, R., Collins, F.H., Venter, J.C. and Hoffman, S.L.	
TITLE	Celera Anopheles gambiae EST project	
JOURNAL	Unpublished (2002)	
COMMENT	Contact: Holt R.A.	

Cellera Genomics
45 W. Gude Dr., Rockville, MD 20850, USA
Tel: 2404533151
Fax: 2404534580
Email: Holter@cellera.com
Plate: NUD010049VS row: J column: 07
Seq primer: M13 Reverse.

FEATURES
source
location/Qualifiers
1..529
/organism="Anopheles gambiae"
/mol_type="mRNA"
/strain="RSP-ST (Reduced susc. to Permethrin - std. chromosome)"
/db_xref="taxon:7165"
/clone="1960049718069"
/dev_stage="Adult"
/lab_host="DHI0b"
/clone_1db="A.Gam.ad.cDNA.blood1"
/note="Vector: pSport1; site 1: SalI; site 2: NotI; whole adult mosquitoes (mixed sex) frozen on liquid nitrogen 24 hours after human blood feeding. cDNA inserts >500 bp cloned directionally into pSport 1. Not 1 site is 3'. Clones available through the Malaria Research and Reference Reagent Resource Center (www.malaria.mr4.org)"

ORIGIN
Alignment Scores:
Pred. No.: 9.76 Length: 529
Score: 75.00 Matches: 22
Percent Similarity: 47.76% Conservative: 10
Best Local Similarity: 32.84% Mismatches: 25
Query Match: 14.59% Indels: 10
DB: 4 Gaps: 3

US-10-089-514-4 (1-103) x BMS95564 (1-529)
Oy 2 ThrGluGlnAenGluLeuGlnArgLeuArgAlaGluLeuAenAlaLeuAenGlyThrLeu 21
Db 377 ACCCGGAGTAATTCACCTTCAGAGGCTG-----CTGGATGTCATTCAGGGGCTTGCT 327
Oy 22 LeuAspThrValAlaArgArgArgIleAspLeuGlyValArgIleAlaArgTyrLysSerArg 41
Db 326 GTGAGCGCGTGC CGGGCGGCTTCGTGGCATGAGG-----CACCTGCGAAGCTC 279
Oy 42 HisGlyValProMetMet-----GlnProGlyArgValSerLeuValLysAspArg 58
Db 278 CACCGAATGCGCGGTATGCTTGTCGTCTGCCGATCGGCGACGACCTACCCGCTTCAC 219
Oy 59 AlaAlaArgTyrAlaAlaAsp 65
Db 218 GTTGAGCGCTGGTTCCAGAG 198

RESULT 11
BH654655/c 644 bp DNA linear GSS 19-FEB-2002
DEFINITION BOMTE10TF_BO_2_3_KB Brassica oleracea genomic clone BOMTE10.
ACCESSION BH654655
VERSION BH654655.1 GI:18712878
KEYWORDS GSS.
SOURCE Brassica oleracea
ORGANISM Brassica oleracea
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Brassica.
1 (bases 1 to 644)
Town,C.D., Van Aken,S., Uterback,T., Koo,H. and Fraser,C.M.
Whole genome shotgun sequencing of Brassica oleracea
unpublished (2001)
Contact: Chris Town
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-3523

PCR Primers
 FORWARD: M13 reverse 17mer at 5' end
 BACKWARD: M13 reverse 17mer at 5' end
 Plate: 002 row: F column: 03
 Seq primer: M13 reverse 17mer at 5' end
 High quality sequence stop: 535.
 Location/Qualifiers

FEATURES
 source
 1..535
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SMY002F03"
 /issue_type="mycelium"
 /cell_line="P6497"
 /dev_stage="mycelium"
 /clone_lib="Agriculture Canada Phytophthora sojae EST
 project_smy"
 /note="Vector: PBK-CMV; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Alignment Scores:
 Pred. No.: 13.6 Length: 535
 Score: 74.00 Matches: 25
 Percent Similarity: 51.79% Conservative: 4
 Best Local Similarity: 44.64% Mismatches: 19
 Query Match: 14.40% Indels: 8
 DB: 7 Gaps: 3

US-10-089-514-4 (1-103) x CF857652 (1-535)

QY 18 AspGlyThrLeuLeuAspThrValArgArgIleAspLeuGlyValArgIleAlaArg 37
 |||||
 Db 264 GACGAGAGAGAGCGGCTCGAGTACGGCGAGCTGCCGAC--GGAGAGAGCTGGCGGC 320
 QY 38 TyrLysSerArgHisGlyValPrometMetGlnProGlyArgValSerLeuValIysAsp 57
 ::|||
 Db 321 CGACGGGGCCGT-----GTCGCCCGTCAGAAACCAAGGCAAGTGGCGCTCGTCTGGAC 374
 QY 58 ArgAlaAlaArgTyrAlaAlaAspHisGlyLeuAspGluSerPheLeu 73
 |||||
 Db 375 -----CTTCTGACACGAGGCTGCTCGAGAGCCACTG 407

Search completed: October 5, 2005, 18:37:47
 Job time : 1132.79 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 19:00:45 ; Search time 4451.25 Seconds
(without alignments)
10548.289 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969

Sequence: 1 atgcgcgcgtcccccgcag.....aagcgagaagacgatga 969

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4708233 seqs, 24227607955 residues

Total number of hits satisfying chosen parameters: 9416466

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb ba: *
2: gb hlg: *
3: gb in: *
4: gb om: *
5: gb cv: *
6: gb pat: *
7: gb ph: *
8: gb pl: *
9: gb pr: *
10: gb ro: *
11: gb ste: *
12: gb sy: *
13: gb un: *
14: gb vi: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	969	100.0	969	6 BD178315	BD178315 Transform
2	969	100.0	969	6 BD093916	BD093916 Transform
3	969	100.0	5251	1 AB116234	AB116234 Streptomy
4	915.4	94.5	14159	1 AF262220	AF262220 Streptomy
5	270.8	27.9	888	6 A48324	A48324 Sequence 2
6	270.8	27.9	888	6 AR198354	AR198354 Sequence
7	270.8	27.9	2888	6 A48323	A48323 Sequence 1
8	270.8	27.9	2888	6 AR198353	AR198353 Sequence
9	270	27.9	4740	1 SF060417	U60417 Streptomyce
10	128.4	13.3	150715	14 AY714813	AY714813 Cercopit
11	128.4	13.3	150715	14 AY714813	AY714813 Cercopit
12	124	12.8	136753	2 AC149968	AC149968 Streptomy
13	122.4	12.6	136753	2 AC149968	AC149968 Streptomy
14	119.6	12.3	1393	11 PM1H12G	AL684264 Penicilli
15	117.4	12.1	156789	14 AF533768	AF533768 Cercopit
16	117.4	12.1	156789	14 AF533768	AF533768 Cercopit
17	116.6	12.0	303450	1 SC0939130	AL939130 Streptomy
18	116.2	12.0	1094	11 PM7G11B	AL685196 Penicilli
19	115.6	12.0	138203	1 AV310323	AV310323 Streptomy

20	115.6	11.9	123580	1 AF263912	AF263912 Streptomy
21	115.6	11.9	125401	6 AX211739	AX211739 Sequence
22	114.8	11.8	65140	6 AX211705	AX211705 Sequence
23	114.4	11.8	1065	11 PM2B12B	AL684695 Actinosyn
24	111.8	11.5	82746	1 AF453501	AF453501 Actinosyn
25	111.2	11.5	321250	1 SC0939111	AL939111 Streptomy
26	110.8	11.4	110000	2 LMFCHR32_06	Continuation (7 of
27	110.4	11.4	22437	1 AY027524	AY027524 Frankia s
28	110.4	11.4	186752	9 AC111200	AC111200 Homo sapi
29	110.2	11.4	186752	9 AC111200	AC111200 Homo sapi
30	109.2	11.3	24045	1 AB070941	AB070941 Streptomy
31	109.2	11.3	325483	1 AP005050	AP005050 Streptomy
32	108.8	11.2	110000	2 LMFCHR36_07	Continuation (8 of
33	108.8	11.2	113193	1 AF357202	AF357202 Streptomy
34	108.8	11.2	113193	6 AX703543	AX703543 Sequence
35	108.6	11.2	296300	1 AP005035	AP005035 Streptomy
36	108.6	11.2	314100	1 SC0939106	AL939106 Streptomy
37	107.4	11.1	969	6 BD178315	BD178315 Transform
38	107.4	11.1	969	6 BD093916	BD093916 Transform
39	107.4	11.1	991	11 PM12H12B	AL684455 Penicilli
40	107.4	11.1	5251	1 AB116234	AB116234 Streptomy
41	107	11.0	299800	1 AP005040	AP005040 Streptomy
42	106.8	11.0	1920	1 AB018799	AB018799 Streptomy
43	106.8	11.0	276800	1 SC0939115	AL939115 Streptomy
44	106.6	11.0	6689	1 AY267372	AY267372 Actinomad
45	106.6	11.0	293050	1 SC0939116	AL939116 Streptomy

ALIGNMENTS

RESULT 1	BD178315	969 bp	DNA	linear	PAT 16-APR-2003
LOCUS	BD178315				
DEFINITION	Transformant producing PF1022 substance and process for producing the same and novel biosynthetic gene.				
ACCESSION	BD178315				
VERSION	BD178315.1	GI:30015580			
KEYWORDS	WO 02077244-A/3.				
SOURCE	Streptomyces venezuelae				
ORGANISM	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomyces				
REFERENCE	1 (bases 1 to 969)				
AUTHORS	Yanai, K., Sumida, N., Watanabe, M., Moriya, T. and Murakami, T.				
TITLE	Transformant producing PF1022 substance and process for producing the same and novel biosynthetic gene				
JOURNAL	Patent: WO 02077244-A 3 03-OCT-2002;				
	MEIJI SEIKA KASEI LTD, KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
COMMENT	TATSUKI MORIYA, TAKESHI MURAKAMI				
OS	Streptomyces venezuelae				
PN	WO 02077244-A/3				
PD	03-OCT-2002				
PR	22-MAR-2002 WO 2002JP002782				
PI	22-MAR-2001 JP 01P 082227				
	KOJI YANAI, NAOMI SUMIDA, MANABU WATANABE, TATSUKI MORIYA, TAKESHI MURAKAMI				
PC	C12N15/61, C12N1/15, C12P17/14, C12N1/15, C12R1:645, (C12P17/14, C12R1:645)				
CC	Transformant producing PF1022 substance and process for CC				
FT	Location/Qualifiers				
	(1)..(966);				
FEATURES	Location/Qualifiers				
source	1..969				
	/organism="Streptomyces venezuelae"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:54571"				
ORIGIN	Query Match				
	100.0%; Score 969; DB 6; Length 969;				

[illegible]

Query Match 100.0%; Score 969; DB 1; Length 5251;
Best Local Similarity 100.0%; Pred. No. 2.7e-66;
Matches 969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
1 ATGAGCGGCTTCCCGCCAGCGTGTGTGCGGCGGAGCGGGGCGTGGGGCGCATGTC 60
2708 ATGAGCGGCTTCCCGCCAGCGTGTGTGCGGCGGAGCGGGGCGTGGGGCGCATGTC 2767
61 GCCGGGCTGTGCGGAGGCGGCGAGCGCGACGCTGCTGTGCACTCTGTACCGCGCG 120
2768 GCCGGGCTGTGCGGAGGCGGCGAGCGCGACGCTGCTGTGCACTCTGTACCGCGCG 2827
121 GGAAGCGCGGACCGCTGCTGTGTGAGGAGAGTCAACCGCGCGGCGCGGCGCGCC 180
2828 GGAAGCGCGGACCGCTGCTGTGTGAGGAGAGTCAACCGCGCGGCGCGGCGCGCC 2887
181 GCCCTCCGGGACGCGGACCTGTCTGTGCGCGGTACAGAGGACGTGGCCCTCAAGGC 240
2888 GCCCTCCGGGACGCGGACCTGTCTGTGCGCGGTACAGAGGACGTGGCCCTCAAGGC 2947
241 GTGGCGCGCGTGAACCGCGCTCATGCGGCGCGGCGCGCTGCTGCGACACCTGTCC 300
2948 GTGGCGCGCGTGAACCGCGCTCATGCGGCGCGGCGCGCTGCTGCGACACCTGTCC 3007
301 CGGACGCGGCAATGGCGCGGAGCTTCGCGGCCCAAGCCCGCGCGTCCAGCACTGG 360
3008 CGGACGCGGCAATGGCGCGGAGCTTCGCGGCCCAAGCCCGCGCGTCCAGCACTGG 3067
361 AACCCGATGTTCCCGCGCGCGCGCGGATGACCGCGCGACCGGTGGCGCGCGTGTACC 420
3068 AACCCGATGTTCCCGCGCGCGCGCGGATGACCGCGCGACCGGTGGCGCGCGTGTACC 3127
421 AGGAGCGGCGCGGCGCGTCAAGCGCTGCTGCGGCTGTGAGAGGCGCGCGGCGGCG 480
3128 AGGAGCGGCGCGGCGCGTCAAGCGCTGCTGCGGCTGTGAGAGGCGCGCGGCGGCG 3187
481 GTACGGCTCAAGCGCGGAGAGCAAGACCGGACGACGCGCGGCGACCCAGCGCTGAC 540
3188 GTACGGCTCAAGCGCGGAGAGCAAGACCGGACGACGCGCGGCGACCCAGCGCTGAC 3247
541 GCGGTGCTCTCTCTTCCGGGCTGCGCGCTGCGCGCGCGTGGCGCGTCCAGCGTCCG 600
3248 GCGGTGCTCTCTCTTCCGGGCTGCGCGCTGCGCGCGCGTGGCGCGTCCAGCGTCCG 3307
601 GCGGCGAGCGGACCGCGCGCGCGCGACAGAGTGTGCTGCGCTGCGCGCGTGTGCTGCG 660
3308 GCGGCGAGCGGACCGCGCGCGCGCGACAGAGTGTGCTGCGCTGCGCGCGTGTGCTGCG 3367
661 GCGAGCGCGCGAGGTGTACGAGGAGCAATCAGCGGTCCAAACCGCGCGCGCGTCCGCG 720
3368 GCGAGCGCGCGAGGTGTACGAGGAGCAATCAGCGGTCCAAACCGCGCGCGCGTCCGCG 3427
721 CGGGCGTGCAGAGGCGCGCTGCGCTGCGCGCGCGTGGTCCGCGACGACCGCGACCGT 780
3428 CGGGCGTGCAGAGGCGCGCTGCGCTGCGCGCGCGTGGTCCGCGACGACCGCGACCGT 3487
781 GCGGAGCGCGCGCGCGCGCGCGACGCGCGCGCATCCCGGAGATTCGACGCGCGCGG 840
3488 GCGGAGCGCGCGCGCGCGCGCGACGCGCGCGCATCCCGGAGATTCGACGCGCGCGG 3547
841 AACCTCGAGCGGCGCTTTCGAGGAACTCCGCGCGGTCAAGGACCGAGCTTCGCGCGCG 900
3548 AACCTCGAGCGGCGCTTTCGAGGAACTCCGCGCGGTCAAGGACCGAGCTTCGCGCGCG 3607
901 CAGAGCACTGTCAGAGAGCTGTTCCGACCTTCCACCGGACCGACGACGAGGCGGAG 960
3608 CAGAGCACTGTCAGAGAGCTGTTCCGACCTTCCACCGGACCGACGAGGCGGAG 3667
961 GACCGATGA 969
3668 GACCGATGA 3676
```

RESULT 4
AF262220
LOCUS
DEFINITION
Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,
partial sequence.
ACCESSION
AF262220
VERSION
AF262220.2
KEYWORDS
GI:14290414
ORGANISM
Streptomyces venezuelae
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomyces
Streptomyces venezuelae
Bacteria; Actinobacteridae; Actinomycetales;
Streptomyces
1 (bases 1 to 14159)
AUTHORS
He, J., Magarvey, N., Piraee, M. and Vining, L.C.
TITLE
The gene cluster for chloramphenicol biosynthesis in Streptomyces
venezuelae ISP5230 includes novel shikimate pathway homologues and
a monomolecular non-ribosomal peptide synthetase gene
JOURNAL
Microbiology 147 (Pt 10), 2817-2829 (2001)
MEDLINE
21461106
PUBMED
11577160
REFERENCE
2 (bases 1 to 14159)
AUTHORS
He, J., Magarvey, N.A. and Vining, L.C.
TITLE
Direct Submision
JOURNAL
Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford
St, Halifax, NS B3H 4J1, Canada
REFERENCE
3 (bases 1 to 14159)
AUTHORS
He, J., Magarvey, N.A. and Vining, L.C.
TITLE
Direct Submision
JOURNAL
Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford
St, Halifax, NS B3H 4J1, Canada
REMARK
Sequence update by submitter
On Jun 5, 2001 this sequence version replaced gi:10716943.
FEATURES
source
1..14159
/organism="Streptomyces venezuelae"
/mol_type="genomic DNA"
/strain="ISP5230"
/db_xref="taxon:54571"
/map="adjacent to pabAB"
233..3511
/function="essential for chloramphenicol biosynthesis"
/codon_start=1
/transl_table=11
/product="CmlG"
/protein_id="AAG21974.1"
/db_xref="GI:10716944"
/translation="MTAGSPGPKTPPRRLAVRDHSFLADGMDVNPCTAFV
GTVFPGRVDPAMREPVARLGLTRPALDLVEVAEDVVTLDPFPGVGGPABLAD
LTVTAHNSRPLRIDEAVRLAYWRHVDGEYLVEKRSYLIVDLADVDLRQIRALRG
DTPGRIIGRYHHHARTVEQYALPAADPAEYTRALGELPVGRKGIPTISASVEML
PLREGVTFDELLSAVTGVLPALGGGLVQYVSRWEFATKGYYVEIKPLVVRGGLP
SDTPVEYFAKTRHAEISGRPTMSDLDPAPAFPSGMRPRAVVSPTTMRPAPDWRV
VYPRARVPEDLKFPLADRSWPGPPLRMQYKRGFLSPGAADVDLGLKORIGASGTP
"

CDS

CDS

3498..6404
/note="non-ribosomal peptide synthetase; contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
/product="CmlH"
/protein_id="AAG21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITLDFRRHGRTRALTAGRTWYELDRVTSALRID
AECPRGRVLVAGHTTAAVYVALAMKSHAVTMTMNGLPADRFEEFARVADALLV
CFERBALVRGERFAGRLAYAGVGPPTDPPAPADGTADPEARSRVAISIFTSGSTGD
PVLVVGHGGLNLCRSRLRLDITPDQVLFALSDAVSEILGLTYAGATLVLP
VVDQASWLGVSVRHLAAGCDLALSPSVYALDAASRIKVFCEALGEGGYDK
AARSRVNAVGTEATVCFSLAELTSTPSIGTVDPGFRAYVRPDSGDHATGAGYK
LYIVGDVALGYAGGSPALENRYGCVDSPPAATDVYSISDGLTYLIGRIDEIKR
LGRVNLAVGSTLSRHIGREVALVRQDATTILVTAAAGEATBSLMARIRDLVFMVE
ADRLVLVALPLTSGKVRDALRELLATLSPAGAGHGDTGDDAELERRVRDLVAVL
GQEIPESTIPDAGSSSLAMIQIVKLSDAYGESEVREAFAMADYDFAPAFRLRIHG
EAVAPAESAVDTLLRRVETERDALRAELPLLRDTRHBPVPGAADGDEVLITGASGF


```

misc_feature
/note="Region: adenyl'ation domain A1"
3708..3746
/note="Region: adenyl'ation domain A2"
3939..3986
/note="Region: adenyl'ation domain A3"
4089..4100
/note="Region: adenyl'ation domain A4"
4350..4370
/note="Region: adenyl'ation domain A5"
4494..4538
/note="Region: adenyl'ation domain A6"
4593..4610
/note="Region: adenyl'ation domain A7"
4647..4706
/note="Region: adenyl'ation domain A8"
4833..4853
/note="Region: adenyl'ation domain A9"
4893..4910
/note="Region: adenyl'ation domain A10"
5043..5075
/note="Region: thiol'ation domain"
5343..5375
/note="Region: NADH binding site"

```

ORIGIN

Query Match	Similarity	94.5%;	Score 915.4;	DB 1;	Length 14159;
Best Local	Similarity	96.6%;	Pred. No. 2.e-62;		
Matches	953;	Conservative	0;	Mismatches	16;
				Indels	18;
				Gaps	1;
Qy	1	ATGAGCGGCTTCC	CCCCGACGATGTCGTGTCGCGCGACGCGGCGCGGTGGGCGGACATGTC	60	
Db	10159	ATGAGCGGCTTCC	CCCCGAGGTGTCGTGTCGCGCGACGCGAGCGGTGGGCGGACATGTC	10218	
Qy	61	GCCGGGCTGCTG	CGGAGCGGAGCGGACGCGACGCTGTGTCGACTTCGTAACCGCGCCG	120	
Db	10219	GCCGGGCTGCTG	CGGAGCGGAGCGGACGCGACGCTGTGTCGACTTCGTAACCGCGCCG	10278	
Qy	121	GGAGGCGCGGAC	CGGCTGCTGTGTGGGGAGGTACCGCGCGGGGGGCCGAATCGCGGGCC	180	
Db	10279	GGAGGCGCGGAC	GCGTCTGTGTGGGGAGGTACCGCGCGGGGGGCCGAATCGCGGGCC	10338	
Qy	181	GCCCTCCGGGAC	CGGACCTGTCCTCGCGGTACACGAGGACGTGACCTCAAGGCC	240	
Db	10339	GCCCTCCGGGAC	CGGACCTGTCCTCGCGGTACACGAGGACGTGACCTCAAGGCC	10398	
Qy	241	GTGGCGCCCTG	ACCCGACTCATGCGGACCGGCGCGCTGCTGCGACACCTGTCCTGTC	300	
Db	10399	GTGGCGCCCTG	ACCCGACTCATGCGGACCGGCGCGCTGCTGCGACACCTGTCCTGTC	10458	
Qy	301	CGGACGGGCAAT	GCGCGGCGGAGCTGCGCGGCCACGCCCGCGGCTCCAGACGTGGGGCTTC	360	
Db	10459	CGGACGGGCAAT	GCGCGGCGGAGCTGCGCGGCCACGCCCGCGGCTCCAGACGTGGGGCTTC	10518	
Qy	361	AACCGGATGTT	CGGCGCGCGCGCGCGGACATACCGGACCGCTGCGCGCGCGGTGTCACC	420	
Db	10519	AACCGGATGTT	CGGCGCGCGCGCGCGGACATACCGGACCGCGCTGCGCGCGCGGTGTCACC	10578	
Qy	421	AGGAGCGGGCG	CGGCGGCTCACCGGCTGTGTCGCGCTGTGAGGGGCGGCGGACAGGCC	480	
Db	10579	AGGAGCGGGCG	CGGCGGCTCACCGGCTGTGTCGCGCTGTGAGGGGCGGCGGACAGGCC	10638	
Qy	481	GTAGGGCTCA	CGGCGGAGAGACACGACCGGACCGCGGGCCACCCAGGCGCTGAGACGAC	540	
Db	10639	GTAGGGCTCA	CGGCGGAGAGACACGACCGGACCGCGGGCCACCCAGGCGCTGAGACGAC	10698	
Qy	541	GCGGCTCTCT	CTTTCGAGCTGCGCCTGCGCCGACTTCGGACTTCGACGTCGCGGCGCTTG	600	

Db	10699	GCCTGATCTCTTCTTCCTTCGAGACTTCGCGCTTCGCGCGCTGACGTCCTCGGAGCTCTG	10758
Qy	601	GCAGCGACGCGACCGCGCGCGCCCAACAGAGTGTGCTTCGCCCTTCGACCCTGATGCTGCAGC	660
Db	10759	CGCGCGACGAGCACCGCGCGCGCCCAACAGAGTGTGCTTCGCCCTTCGAGCGCTGTGTGCTCGCGC	10818
Qy	661	GGCAGCGCCCGAGGCTGTAACGGGGACATTCACGCGGTCCAACTCCCGGCGGCGCTTCGCGCGCGC	720
Db	10819	GGCAGCGCCCGAGGCTGTAACGGGGACATTCACGCGGTCCAACTCCCGGCGGCGCTTCGCGCGCGC	10878
Qy	721	CGGCGCGCTTCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGC-----	768
Db	10879	CGGCGCGCTTCGCGCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGCTTCGCGCGC	10938
Qy	769	-----GACCCGGAACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	822
Db	10939	GCCGAGGACCCCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	10998
Qy	823	GGATGCGACGCGCGCGCGGAACTTCGACGCGCTTCGCGGGAATCTCGCGCGGCTCATGGA	882
Db	10999	GGATGCGACGCGCGCGGGAACCTTCGACGCGCTTCGCGGGAATCTCGCGCGGCTCATGGA	11058
Qy	883	CCGAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	942
Db	11059	CCGAGGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	11118
Qy	943	GACGACGAGCGCGAGGACCGCATTA	969
Db	11119	GACGACGAGCGCGAGGACCGCATTA	11145

RESULT 5

LOCUS	A48324	888 bp	DNA	linear	PAT 07-MAR-1997
DEFINITION	Sequence 2 from Patent WO9601901.				
ACCESSION	A48324				
VERSION	A48324.1	GI:2302117			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS	Streptomyces pristinaespiralis ; Streptomyces pristinaespiralis ; Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces. 1 (bases 1 to 888) Blanc, V., Thibaut, D., Bamae-Jacques, N., Blanche, F., Crouzet, J., Barriere, R., Debussche, L., Famechon, A., Paris, J. and Dutruc-Rosset, G. STREPTOMYCEINASES AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS Patent: WO 9601901-A 2 25-JAN-1996; RHONE-POULENC RORER SA (FR) Other publication AU 2891295 960209 Other publication FR 2722210 960112. Location/Qualifiers 1..888 /organism="Streptomyces pristinaespiralis" /mol_type="unassigned DNA" /db_xref="taxon:38300"				
FEATURES					
source					
ORIGIN					
Query Match	27.9%;	Score 270.8;	DB 6;	Length 888;	
Best Local Similarity	58.8%;	Pred. No. 1.5e-12;			
Matches	486;	Conservative 0;	Mismatches 337;	Indels 3;	Gaps 1;
Qy	10	TTCCCCCGCAGCGTCGTCGTCGCGCGCAGCGCGCGCGCGCGCAAGTT---	CGCCGGG	66	
Db	19	TTTCGGGGGTGTGTGTGTGTGTGGGGGGGGCGGTGTGGTGGCCGCACTTTCAGCCACTGG		78	
Qy	67	CTGCTGCGGGAGAGGTGGGGCAGCGCCGACGCTGTCGTGCACTCTGTCAGCCGCGCGGAGCG		126	
Db	79	CTGTGTGCTTTCGGGGGTGGCGCGTACCTGTGCTGACGTGACCGGGGCGGTCGCGCGGAC		138	
Qy	127	CCGAGCGCTTGCTGTGTGGCGCAGTCGACCGCGCGCGGGGCGGAAGTCTGCGGCGGCTTC		186	

Db	139	GGGGGGGCGGGGTGTGGCCGGTATGTGGCGCGGCCGGGGCCGGAGACGGTGCGGCGCTG	198		
OY	187	CGGACGCGGACCTTCGTCTCTGCTCGCCCGTACAGAGACGTGGCCCTTCAAGGCGCTGGCG	246		
Db	199	GCGGCGGGCGGACGTGTGTGTCTGTGGCGGGTCCGGAGCGAGTGGCTGGAGAGGCGGGTGGAG	258		
OY	247	CCCGGACCCCGGCTATGCGGCGCGGGCGCGCTGTGGCGACACCTGTCCGTCCGGACG	306		
Db	259	GTGCTGGCGGGGGTATGTGGCGCCCGGTGGGTGTGTGGCGAACCTTGTTCGTTCAAGAGC	318		
OY	307	GGCATAGGCGCGGAGCTGCGCGGCCACGACCCCGCGCGTTCAGACAGTGGGCTTCAACCG	366		
Db	319	CGGATCGCGCGCGGCGGCTGCGTAGGCGCGGCCGGGCTGTCAAGCGGTTGGGCTGAACCG	378		
OY	367	ATGTTTCGCCCGCGCGCGCGCATGACCGGCGGACCCGTTGGCCCGCTGTGTCAACGAGAC	426		
Db	379	ATGTTTCGCCCGCTCGCTGGGTCTTTCAGGGGCGGCGCGGTGGCGGGCGGTGTGATCACGAC	438		
OY	427	GGGCGCGGGCGTCACGGGCCCTGCTGTGGGCTTCGTTCAGAGGCGCGCGGGCGGCGCTTACGG	486		
Db	439	GGGCGCGGGTGTGCGGGCCCTGTGTGAGCTGTGTGGCGGGGTGGGGGGCCCGGTTGTGGAG	498		
OY	487	CTCACGGCGGAGAGACACGACCGGAGCGACGCGCGCACCGGACGCTTGAACGCAACGCGTG	546		
Db	499	ATGCGCGGGCGCGCGGCGACGACGACTGACCGCGCGGCGACAGAGCGCGCACGATGTCGCG	558		
OY	547	CTCCTCTCTTTCGAGGCTGCGCTCGCCCGCGCTCGGCGTGAACGTGCGGCGCCTTGGCGGCG	606		
Db	559	GTGCTGGCGCTTTCGAGCTGGGCTCTGGGTGAGCTGTGCGGTGAGACGTGGGGCGCGTGGGAGC	618		
OY	607	ACGGACACGCGCGCCCAACAGGTGTGTCTGCGCCCTTCGAGCGCGCTGTGTCTGGCGGACG	666		
Db	619	AGTGCCTCCGCGCGCGCATCTGTGCGATGTCTGGGCTGTGCGCCCGGATGCGCGGCGGACG	678		
OY	667	CCCGAGGTGTACGCGGAGCATCGAGCGGTTCGAACCCCGCGGCGGCGCTCCGCGCGCGGACG	726		
Db	679	CCGAGGATGTATTTCAGATATCCAGGCGCGCAACCCCGGCGCGCGCGCGCGCGCGGACGCG	738		
OY	727	CTCGCCGAGGCGCTTGCCTCTTCTCGCGCGCTGTGTGCGCGACGACCCCGACCGTGCACG	786		
Db	739	CTGGGCGCGCGCTGTGTGCGGCTGTGGGCGGCGCGCTGAGAGGGGCGACGAGAGACGTTTC	798		
OY	787	GCCCCCGGGGCGCGGACGCGCCCGCGGCGCATCCCGGGGATCGACG	832		
Db	799	GCCGCGCTGTTTGCAGAACTGCGCGGTGTGCTGGGCGAGCACGCGG	844		
RESULT 6					
LOCUS	AR198354	888 bp	DNA linear PAT 20-APR-2002		
DEFINITION	Sequence 2 from patent US 6352839.				
ACCESSION	AR198354				
VERSION	AR198354.1	GI:20248203			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				
REFERENCE	Unclassified.				
AUTHORS	1 (bases 1 to 888)				
TITLE	Blanc,V., Thibaut,D., Bamae-Jacques,N., Blanche,F., Crouzet,J.,				
	Barriere,J.-C., Debussche,L., Fanechon,A., Paris,J.-M. and				
	Durruc-Rosset,G.				
	Streptogramins for preparing same by mutasynthesis				
	Patent:US 6352839-A 2 05-MAR-2002;				
JOURNAL	Location/Qualifiers				
FEATURES	1..888				
source					

ORIGIN	/mol_type=	unassigned DNA
Query Match	27.9%	Score 270.8; DB 6; Length 888;
Best Local Similarity	58.8%	Pred. No. 1.5e-12;
Matches 486; Conservative	0;	Mismatches 337; Indels 3; Gaps 1;

[illegible]

RESULT 7					
A48323/c					
LOCUS	A48323	2888 bp	DNA	linear	PAT 07-MAR-1997
DEFINITION	Sequence 1 from Patent WO9601901.				
ACCESSION	A48323				
VERSION	A48323.1	GI:2302116			
KEYWORDS	.				
SOURCE	Streptomyces pristinaespiralis				
ORGANISM	Streptomyces pristinaespiralis				
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
	Streptomycinae; Streptomycetaceae; Streptomyces.				
REFERENCE	1 (bases 1 to 2888)				
AUTHORS	Blanc, V., Thibaut, D., Bamae-Jacques, N., Blanche, F., Crouzet, J., Barriere, J., Debussche, L., Farnecion, A., Patis, J. and Dutric-Rosset, G.				
TITLE	STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS				

JOURNAL

Patent: WO 9601901-A 1 25-JAN-1996;
RHONE POULENC RORER SA (FR)
Other publication AU 2891295 960209
Other publication FR 2722210 960112.

COMMENT

FEATURES

source
1. 2888
/organism="Streptomyces pristinaespiralis"
/mol_type="unassigned DNA"
/db_xref="taxon:38300"

ORIGIN

Query Match 27.9%; Score 270.8; DB 6; Length 2888;
Best Local Similarity 58.8%; Pred. No. 8.8e-13;
Matches 486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

10 TTCCTCCGAGCGCTCGTCGCGCGCAGCGGGCGGTGGCGGACATGTT--CGCCGG 66
1818 TTCGGGCGTTGTGTGTGTGTGGCGGGCCGTCGCGGCGCATGTTCAAGCCACTGG 1759
67 CTGCTGCGGGAGCGGGCAGCCCGACGCTGTCGACCTGTAACCGCCGCGGACGG 126
1758 CTGTGTGTTTCGGGGGTGGCGGTGACCTGAGCTGGACCGGGCGGTGCGCGGAC 1699
127 CCGGACGCTGCTGTCGTGCGGACGTCACCGCGCGGGCCCGAATCGCGCGCTTC 186
1698 GGGGTGCGGGTGTGGCGGGTGAATGTCGGCGCGCGGGCCGAGAGCGGTTCGCGCGCTG 1639
187 CCGGACGCGGACCTGTCCTGTCGCGGTCACAGAGACGTGGCCCTCAAGCGCTGGCG 246
1638 GCGGCGCGGACGTCGTGTGTGTGTGGCGGTCCGAGCGGGCGGTGGAGGCGGTGGAG 1579
247 CCGGTACCGCGGCTCATATGCGCGCGGGCGGTGTCGCGGACACCTGTTCCTGCGGAC 306
1578 GTGCTGCGGGGTGTGATGCGGCGCGGTGTGTCGCGGACACTTGTGTGTCAGAGCG 1519
307 GGCATGCGCGCGGACCTGCGGCGCCACGCGCGCGCGCTCCAGACGTGGGCTCAACCG 366
1518 CGGATCGCGCGGCGCGGTGCGGTGAGCGCGCGCGGTGAGCGGTGGGCTGAACCG 1459
367 ATGTTGCGCGCGCGCGCGGATGACCGCGGACCGGTGGCGCGGTGTGTCACAGGAC 426
1458 ATGTTGCGCGCGCGCTGCGGTGTCAGGGGCGCGCGGTGGCGGTGTGTCACGAC 1399
427 GGGCGCGGCGTCAACGCGCGTGTGCGGTGTGTCAGAGGCGCGCGCGGACCGGTACCG 486
1398 GGGCGCGGTGTGCGGCGCTGTGTGAGCTGTGTGCGGTGTGCGGCGCGGTGTGTGAG 1339
487 CTCACGCGGAGAGACGACGCGGACGCGGCGGACCGCGCGCTGACGCGCGGTG 546
1338 ATGCGCGCGCGCGGCGGACGAGCTGACCGCGCGGAGAGCGCGCGCATGCGCG 1279
547 CTCTCTCTCTTGGGCTCGCGCTCGCGCGCTGCGGTGTCAGCTCCGAGCGCTGGCGCG 606
1278 GTGCTGCGCTTTCGGGCTGCGGCTGTGTGAGCTGTGCGGTGACGTGGGCGCGCTGGGAC 1219
607 ACAGGACCGCGCGCGCGGACGAGTGTGCTGCGCTCTGCGCGGTGTGTGCTGCGCGGAC 666
1218 AGTGCCTCGCGCGCGGACATGCGGATGTGCGGTGTGCGCGGTGTGCGCGGAGCG 1159
667 CCGGAGGTGTACGAGGACATCAGCGGTGTCAACCGCGCGGCGGTGTCGCGCGCGGAG 726
1158 CCGGAGGTGTATTTCAGATTCAGGCGCGGACCGCGCGCGCGCGCGCGCGGAGCG 1099
727 CTGCGGAGGCGCTGCGCTCTTTCGCGCGCGGTGTGCGGACGACCGCGGACCGTGCAC 786
1098 CTGGGCGCGCGGCTGTGTGCGGTGTGGGACGAGCGGTGTGAGAGGAGCGACGATTC 1039
787 GCGCGCGGCGCGCGGACCGCGCGCGGACCGCGCGGAGTGTGCGCGCGCGGAGCG 832
1038 GCGCGCTGTGTGCGCGGACGCGGCGGTGTGCGGCGGAGCGCGCG 993

RESULT 8

AR198353/c

AR198353 2888 bp DNA linear PAT 20-APR-2002
Sequence 1 from patent US 6352839.

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITILE

JOURNAL

FEATURES

source

ORIGIN

Query Match 27.9%; Score 270.8; DB 6; Length 2888;
Best Local Similarity 58.8%; Pred. No. 8.8e-13;
Matches 486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

10 TTCCTCCGAGCGCTCGTCGCGCGCAGCGGGCGGTGGCGGACATGTT--CGCCGG 66
1818 TTCGGGCGTTGTGTGTGTGTGGCGGGCCGTCGCGGCGCATGTTCAAGCCACTGG 1759
67 CTGCTGCGGGAGCGGGCAGCCCGACGCTGTCGACCTGTAACCGCCGCGGACGG 126
1758 CTGTGTGTTTCGGGGGTGGCGGTGACCTGAGCTGGACCGGGCGGTGCGCGGAC 1699
127 CCGGACGCTGCTGTCGTGCGGTCACAGAGACGTGGCCCTCAAGCGCTGGCG 186
1698 GGGGTGCGGGTGTGATGCGGCGCGGTGTGTCGCGGACCGGGCGGTGGAGGCGGTGGAG 1639
187 CCGGACGCGGACCTGTCCTGTCGCGGTCACAGAGAGTGGCCCTCAAGCGGTGGCG 246
1638 GCGGCGCGGACGTCGTGTGTGTGTGGCGGTGTCGCGGACCGGTGGAGGCGGTGGAG 1579
247 CCGGTACCGCGGCTCATATGCGCGCGGGCGGTGTCGCGGACACCTGTTCCTGCGGAC 306
1578 GTGCTGCGGGGTGTGATGCGGCGCGGTGTGTCGCGGACACTTGTGTGTCAGAGCG 1519
307 GGCATGCGCGGACCTGCGGCGCCACGCGCGCGGTGTCAGAGCGGTGTGTCACGAC 366
1518 CGGATCGCGGCGCGCTGCGGTGAGCGCGCGCGGTGACGCGGTGGGCTGAACCG 1459
367 ATGTTGCGCGCGCGCGGATGACCGCGGACCGGTGGCGCGGTGTGTCACAGGAC 426
1458 ATGTTGCGCGCGCGCTGCGGTGTCAGGGGCGCGCGGTGGCGGTGTGTCACGAC 1399
427 GGGCGCGGCGTCAACGCGCGTGTGCGGTGTGTCAGAGGCGCGCGCGGACCGGTACCG 486
1398 GGGCGCGGTGTGCGGCGCTGTGTGAGCTGTGTGCGGTGTGCGGCGCGGTGTGTGAG 1339
487 CTCACGCGGAGAGACGACGCGGACGCGGCGGACCGCGCGCTGACGCGCGGTG 546
1338 ATGCGCGCGCGCGGCGGACGAGCTGACCGCGCGGAGAGCGCGCGCATGCGCG 1279
547 CTCTCTCTCTTGGGCTCGCGCTCGCGCGCTGCGGTGTCAGCTCCGAGCGCTGGCGCG 606
1278 GTGCTGCGCTTTCGGGCTGCGGCTGTGTGAGCTGTGCGGTGACGTGGGCGCGCTGGGAC 1219
607 ACAGGACCGCGCGCGCGGACGAGTGTGCTGCGCTCTGCGCGGTGTGTGCTGCGCGGAC 666
1218 AGTGCCTCGCGCGCGGACATGCGGATGTGCGGTGTGCGCGGTGTGCGCGGAGCG 1159
667 CCGGAGGTGTACGAGGACATCAGCGGTGTCAACCGCGCGGCGGTGTCGCGCGCGGAG 726
1158 CCGGAGGTGTATTTCAGATTCAGGCGCGGACCGCGCGCGGAGTGTGCGCGCGGAGCG 1099

[illegible][illegible]

QY 427 GGGCGGGGCGTCACGGGCGCTGCGGGGCTGTCGAGGGCGCGCGCGGCGGACCGCGTACGG 486

Db 2938 GGGCGGGGCTGTCGGGGCCCTGCTGGAGCTGTGGCCGGGTGGGGGCGCCCGGTGTGGAG 2879797

QY 487 CTCACGGCGGAGAGACGACGACCGACGACCGCGGCGCACCCAGGCGCTTACGACGACGCGGTG 546

Db 2878 ATGCCGCGCGCGGCGGACGACGACGTGACCGCGCGCGGCGACGAGGCGCCACGATGCCGCG 28119

QY 547 CTCCTCTCTTTCGGGCTGCGCTGCGCGCGCTCGCGCGTTCGACGTCGCGGCGCTTGGCGGCG 606

Db 2818 GTGCTGGGCTTTCGGGCTGCGGCTGCGGCTGCGGTAGCTGTTCGGTGGACGTGGGGCGCGCTGCGGGAC 27598

QY 607 ACGGACCGCGCGCGCCACGAGGTGCTGTCGCGCGCTCTGAGCGCGGTGCTCTGCGGGCGAC 666

Db 2758 AGTGGCGCGCGCGCGCATCTGGCGATGCTGGGCTGCTGGCGCCGGATGCGCGGAGGACG 26999

QY 667 CCGGAGGTGTACGGGAGCATCCAGCGGTCACAAAGCCCGGGGCGGTCGCGCGCGCGGACG 726

Db 2698 CCGGAGGTGTATTTTGAATCATCAGGCGCGCAAGCCCGGCGGCGCGCGCGCGGAGGCG 26399

QY 727 CTGCGCGAGGCGCTGCGGCTCTCTTGGCGCGCGCTGTGCGCGACGACCGGACCTGTGCCAG 786

Db 2638 CTGGGCGCGCGCGCTGTGTGCGGCTGCGGCGAGGCGCGCTCGAGAGGGGAGACGAGAGACGTTC 257979

QY 787 GCGCCCGGCGCGCGGACGCGCGCG 812

Db 2578 GCGCGCTGTTCGCGGAACTGCGCGG 2553

polyA_signal
 polyA_signal
 polyA_signal
 gene
 CDS

RAAAVAVRPPAPSDPILLIDSPSPRRPPAVSGRPVAVYAPRPAPMPRPAPARAPR
ALTOAQAQTOQASQARQAQALLAQALQALGRAPPRAPRPOHNTQTPRPAQQTOT
PLQAARQTAQTAQTAQTAQQAQAQAQQTQTQTARPERQTQAQTAQTAQQTQTQ
RRPAPAGAGAGSGSRGPKRASLPAPDAAPRPAQLPPLAAPAPAPPPAPSS
GPGSAPAPAPAPAPAPAPAPAGSLGPRPAPRGCRKAKTHTADADAAPASGTRY
PLSGVSVAVAMPYLNKYTQTGCLPLIDMETGAIGAVYLWGDQCNARCLADEPQW
ARRSPRPPAAPPCVSPPEYFGDPBAHGLWMTPVQGMPLFGQALLGGRSFHSLDSNHPMT
PAPSDP

4856. .4861
complement (4879. .4884)
complement (8315. .8320)
9469. .10167
/gene="U11"
/gene="U11"
9469. .10167
/gene="U11"
/note="complexes with gH; involved in membrane fusion"
/codon_start=-1
/product="viral glycoprotein L"
/protein_id="AAU88066.1"
/db_xref="GI:5284375"
/translation="NAVAVRGLVPLASLPFCSFNNILSTPDSSTTEMLTQAAVAVV
ADVLRQGLPADQVTVRVAAPRSATDVARIGIPLRHQPGDYLTMDSQAAVAVV
NPLPTAGCLELDSHALFPDNLDELTRALYKEVRLVAVSRGASDITEVPECVSS
KRRSRRCRGRTGPGFPEBRPLRSLPASPADAQPFSPAATSSAPPRARRKPPR
KRRSRRTESRRR"
10235. .10990
/gene="U12"
10235. .10990

RESULT 10
AY714813

gene
CDS

LOCUS	AY114813	150715 bp	DNA	linear	VRL 04-OCT-2004
DEFINITION	Cercopithecine herpesvirus 2, complete genome.				
ACCESSION	AY114813				
VERSION	AY114813.1	GI:52843273			
KEYWORDS					
SOURCE	Cercopithecine herpesvirus 2				
ORGANISM	Cercopithecine herpesvirus 2				
	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Alphaherpesvirinae.				
REFERENCE	1 (bases 1 to 150715)				
AUTHORS	Tyler,S.D., Peters,G.A. and Severini,A.				
TITLE	Complete genome sequence of Cercopithecine herpesvirus 2 (SAB) and comparison with other simlexviruses				
	Virology (2004) In press				
JOURNAL	2 (bases 1 to 150715)				
REFERENCE	Tyler,S.D., Peters,G.A. and Severini,A.				
AUTHORS	Direct Submission				
TITLE	Submitted (11-AUG-2004) Health Canada, National Microbiology				
JOURNAL	Laboratory, 1015 Arlington St., Winnipeg, Manitoba R3E 3R2, Canada				
FEATURES	Location/Qualifiers				

polyA_signal
gene
CDS

```

/transl_except=(pos:10235..10237,aa:Met)
/product="uracil-DNA glycosylase"
/protein_id="AAU88067.1"
/db_xref="GI:55843376"
/translation="MTFSPDAPALGLDDTPAGRPSPLMPEERERFLVGDAMPFLLE
PELAPALRLAAEXERERAREEVLPPREDSWMTFCMDEYRVILIGDDPHAPQDQ
AHGLAFSRVGRPAPRPSLNLNTIYVANCCEPRTMSHGCLKEKARAGVLLNTLTLYR
RGEPSHAKIIGMDRVGSVVRRLASRPFLVFMVLWGAHAKQSALRPDPRVHRLVLTAPR
SPLSRVPGSGCRHFCFLANQYLERSLAPIDWST"
11076..11081
11130..11183
/gene="UL3"
11130..11183

```

```

source
1..150715
/organism="Cercopithecine herpesvirus 2"
/mol_type="genomic DNA"
/db_xref="taxon:10317"
/note="SAB"
1..236
/misc_feature
/note="a sequence"
237..9207
/feature="terminal copy of long repeat region"
2178..4669
/gene="RL2"
join(2178..2225,2448..3020,3398..4669)
CDS

```

polYA_signal
polYA_signal
gene
CDS

/translat10="MSA"VADPVSSTLAVLSMGFAFHDGPPAAAEKSPADAPAP
 PUGHVPRPADRDAPAPGURAPADPVAPEPTMWSVSDYLGROLDTTRKIDRLSLV
 NLAICTKTSSESGTAGARAPAAHGGPREPPSHSLQMFVLÇPDPARVQDQLAA
 VLAASEPKKYYTTRSDGRISPAIPVFVHEFVSDAPVHARDNVIAEVPAPRRISTRLA
 DPASEPQ"
 11809. .11814
 complement (11832. .11837)
 complement (11897. .12511)
 /gene="U74"
 complement (11897. .12511)

```

/gene="RI2"
/note="multifunctional regulatory protein"
/codon_start=1
/product="immediate early protein ICP0"
/protein_id="PAU88065.1"
/db_xref="GI:52843274"
/c translation="MEPSRGIAGPGRSSAQAAGALDEELMLSRFFPSSDSDSD
AAPDCCGDRBPASDADTDAIECGPAGSGPRAPSPPPRVCATIRIETQIC
VAEPCLHRFCIPCKITMLPKRNSCPICNAVYVITLVKXKDSYSTIPVINDRPTCE
AEEAVRGTAVDITWITRLPGEALPASVITIGKRVRLSPARRGCPAPGGAARAP

```

gene

```

nuclear_bodies"
/codon_start=1
/product="UD4"
/protein_id="AAU8606.1"
/db_xref="GI:52843278"
/translation="MSASASATIIYSLHGAATSCACVLPDAEOVVCAPESGRIALAS
RGCLRHLDSLGAIVYRPTGLIVMDCRCEFCAYRFGREBRLERLMDITGLAY
PFDSVNSTRESASRSPAGIATIVYMGEDSIITYITTVYGSPEEGEGAGAPRLPPPGD
APDAASPVAVSPASASTAADALIVEMKEIQISLPTLGGPCDA"
complement(12536..15163)

```

gene="U15"
complement (12536, .15163)
/gene="U15"
/codon_start=1
/product="component of helicase-primase complex"
/protein_id="AAU88070.1"
/db_xref="GI:52843279"
/translation="MTTVGESAGDASAPPEQPPVPSARFRDEAFINFTSMHIQPI
IBERVLAGHRLPARPRPOMFRVDALETADLPRLREFPAVYLITNGSGKSTC
VOTLNEVLDCVVTGATRIAQNNYVLSGFLSRPINTIEHFGRNNGVOAGHP
VYLSNPASVEDLORDLTYMEVITADIRBAVAGGSDGGEFRALALERTMGL
AGSLURLAFTHGALPAFTRSITVIVIDEAGLLRLTAVTYCMNTAALHYTPYA
AGLRPLVCSEPTQASLESTIEHOKLCSIKOSNVILYLCNKRLTYVRLPESW
AIFINNKRCVEHEFNIMKALEXGLPLTEHHQPDVRFVPSFITNPANLPQMTLF
SSHKEVSAMARLHAHLKVAREGEFVFTLPVLTFSVTKAFEEYRELTPALTKVM
IANAGRIITNYSOSODODAGRVFEVHSKOOLVANDVTYVINSQIATVTRLKVF
GGCTFRFEAVLRDSDPVKTOGEMLVEEPRLASLITGGLIHFPNLLRPLDAK
RARAYERMALTTTELLALGABEGABGAAPPAAGRAADFYLKARAAAGDEFEPPD
LDVFAVLSSELDLFTYCHVPEQPTTAVHAQFELIKRAFLGRYHILRELFGFAFE
GAFFSTYVDNVTFRGCEWIMGSRGGLMSVALQDTGYTLTGFTYARVYAFABVRBH
AAGLAELLEESPLPYVLRDGHFMSVNVTVSEVSEARLMAATNADYGISM
LMTITRSOGLSLDKVAICTPGNLRILNSAVYMSRTSSEFLRMNLNPLRERHERDD
VTSGLTSLARDEVLIVY"
15162, .17198
/gene="U16"
15162, .17198
/gene="U16"
/note="role in DNA cleavage/packaging"
/codon_start=1
/product="capsid protein"
/protein_id="AAU88071.1"
/db_xref="GI:52843280"
/translation="MAERHPRLRPAMPSPDGGPRPADGWVQAHPTPTMLFREALS
GELGYEGGVVNVVRSSEATKLOAAVFGALNATNARLEADMDHRAVRAARLEP
RLVRRAGAPDAVAAVAREVPTWRGALRASLLDARGADCPACCGGGSGYGR
IDMMLGLGVYVNRPRRBSGVTOLEAFIAGHALPDILTVGAVRRAAPFTALARA
PBARLADYDVCLFTYHPRGEWVRDPAAGRGGEVCVIMPLMAGDRLLPDSVPQRL
SHEITACRALREHARCVRNASVAVLGRKNDGDRGAGAAVRAALGEBDASRA
GSAASRLVLIIMKGMRIIDINDTVRAVLDGAGHLMDSATVDALGFGAAGAG
RAAGADPAPGRQOOLQAFQTSVNNINGMLGYINNLFGTIERLRENTDILATOLR
DRLELRBARGLDRDEORADLGREGGGAAPDPBPADYDIDYSKMDNTYVAN
SRHOYVPSYQDLERLSRLMEHELLRCKTISRHTNOCBETIPISSGALAFVAPY
FADLRAPILPGALITGPBGVLGSEELMEAI PKTTRQTYLTDLALFVADVRRAARR
LPAAEADPGRIRRSRSPVDAGRPGARGEDAVWGHDHGRPELRGRGGRP"
17149, .18039
/gene="U17"
17149, .18039
/gene="U17"
/note="role in DNA cleavage/packaging"
/codon_start=1
/product="capsid protein"
/protein_id="AAU88072.1"
/db_xref="GI:52843281"
/translation="MAPSGSEBETAAALROAVAGDRDLVGVAALISQOALIRNACE
APVAAGRDVAFATGVSRVDVTPDRRLRALDGRDPAATASDYPARCGOAAHNGF
ARVAVLAGEHDVHSLCVPLVLPYHRLSVRPERLEDFELACLMLYENCPRAATASA
FYKISAMLAAGRTSPDRVRCLLRSCOMLNTFMVHYVPPFDRVYVLPHMCAAR
YLDHNPPLITLAFRAASRVFRLPVRGTNLECVANYPDGLSGGMAADPRAGLA
AMWASDAKPTQTAALFYRYV"
18062, .18067
complement (18205, .18210)
/gene="U18"
complement (18213, .20489)
complement (18213, .20489)
/gene="U18"
/codon_start=1
/product="component of helicase-primase complex"
/protein_id="AAU88073.1"
/db_xref="GI:52843282"
/translation="MGEDGLAAEESVCATVLYAAWLPRTDRILALYLICORDARG
DAYARPAEVAVASSELOAFYGPGEISAGGVAARAASPAAPLETLNDPALMRALY
ASALAALREBVGAPLFPARVGRDPRGTGVARI EAPVGRVAPPGALAGVDAQVAT
DPLAALAAVAAAPARLAWARLAAIODTPOCASASIALAVRIDGTARFAREYTTLTFP

PRKSGANGDVFEVREANLPRGHQAQATYAVLVPRGYEYAVRABGSPAPLVALFR
QNHILTVHAAPGALPPIFAFLGPEFEPREGADVHFEVLGPPGPTIKLKPVGAAPVAA
LTHAHELAGAMPAGARALAPARLMAAAGAPGLSPALRRAPRTKLPVAVPVG
PYCARFRPGLFARLAAALCGLEPRAADPDPARGRCRRLAADRPAARVAA
IALABARECPAFARLVGVAAVGLRIEAAAGVDVLAVCGNEGSLMVLVAVDPCA
AGPAVERPARALBGNVAGALAGGLLREPPRBLTEGITYTHAVINSGCATMENA
RTDADRLBGFQORGPAYIAAAATRRALRGAVAAVABEGAKAKLALBEALDGLVLE
ATARLDEWMAVAGPSDPGDTPPAALRGALLDSEHRSRVVRCVCGPDGGSVAV
PIDLYRPLVPEPIDCAHLRLDVLAEBEAVLRGALLMGEDGGRFAYPEERFLPLFD
PGGAAGHAVARSGSRKA"
complement (20544, .23183)
/gene="U19"
13.3%; Score 128.4; DB 14; Length 150715;
Best Local Similarity 47.5%; Pred. No. 0.012;
Matches 381; Conservative 0; Mismatches 421; Indels 0; Gaps 0;
Query Match
6 CGGCTTCCCCCGCAGCGCTGCTGTCGGCGGACAGCGGGGCGGTGGCGGCGATGTCGCCG 65
147992 CGCGCGCGCGCGCGCGCTGCTGTCGTGACCGCGCGCCGAGAGTGGCCGCGCA 148051
66 GCTGCTGGGGAGGCGGGGAGCGGACCGCTCGTGTGACCTGTACCGCGCGGGAGG 125
148052 CGGCGCCGCTGCTGCGCGCGCGCGGACGTACCTGCGCTGCGCGCGCGCGCGCA 148111
126 GCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 185
148112 GTGCGACGTGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148171
186 CCGGAGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
148172 CGTCTTGGGGCGCGCGCTTGTGCGCGCTGCAAGCGCGCTGCGCGCGCTTACCCCA 148231
246 GCGCGGAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305
148232 CGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 148291
306 GGGCATGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 365
148292 GGGGCGCGGACACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 148351
366 GATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
148352 GCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148411
426 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 485
148412 CTTCCGCGCGCGGAGGCGCACTCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148471
486 GCTCAAGCGCGAGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 545
148472 GCTGCGCGCGCGCTACCTGCGCGCTGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148531
546 GCTCTCTCTCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 605
148532 GCGCGCGCGCGCTGCGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148591
606 GACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
148592 GCGGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148651
666 CCCCGAGGTGTACGAGGACATCCAGCGATCCAAACCCCGAGGCGGCTCCCGCGCGCGCGCG 725
148652 CTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148711
726 GCTCGCGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 785
148712 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148771
786 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
148772 GAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 148793

RESULT 11
AY714813/c 150715 bp DNA linear VRL 04-OCT-2004
LOCUS AY714813
DEFINITION Cercopithecine herpesvirus 2, complete genome.
ACCESSION AY714813
VERSION AY714813.1 GI:52843273
KEYWORDS
SOURCE
ORGANISM
Cercopithecine herpesvirus 2
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae.
1 (bases 1 to 150715)
REFERENCE
AUTHORS Tyler,S.D., Peters,G.A. and Severini,A.
TITLE Complete genome sequence of Cercopithecine herpesvirus 2 (SAB) and
comparison with other Simpliciviruses
JOURNAL Virology (2004) In press
AUTHORS Tyler,S.D., Peters,G.A. and Severini,A.
TITLE Direct Submission
JOURNAL Submitted (11-AUG-2004) Health Canada, National Microbiology
Laboratory, 1015 Arlington St., Winnipeg, Manitoba R3E 3R2, Canada
FEATURES
source
1. 150715
/organism="Cercopithecine herpesvirus 2"
/mol_type="genomic DNA"
/db_xref="taxon:10317"
/note="SAB"
misc_feature
1. 236
/note="a sequence"
237..9207
/note="terminal copy of long repeat region"
2178..4669
/gene="RL2"
/join(2178..2225,2448..3020,3398..4669)
/gene="RL2"
/note="multifunctional regulatory protein"
/codon_start=1
/product="immediate early protein ICP0"
/protein_id="AAU88065.1"
/db_xref="GI:52843274"
/translation="MEPSRGLGAPGRSSAQAADALDELLMLSRFFPSSDDSD
VDDDDGGRRPADADTDAELCEPGSRGPRASPPRPRECAVCTERIDETOLC
AARPCLRFCIPCLKTWLPMBNSCPICNAVAVLIVGKDSYSTIPVNDRTAE
AERAVRAGTVDTWTHTLPEGAAPSVTLGGRTVRLSPAMGQAPRGAAPAB
RAARVAPRPPADSPILIAOSPPASPPRPAVAGPPVAPVAPPRAPRAAPAP
ALTOAQAOTQAOQAQAALQAALQAALGAPRPAAPQAHATTOTPRQAOTOT
PTQARAOTQAOQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQAQA
RKPASGAGSGSRGPKRASLPAPDAPARAPOLPAPALAAAPRAPAPASASA
PRGSAPPRPAPAPRPAARAGSSLGRRPARGRKCARKTHYDADRAPASGPTTYL
PISGVSSVAMAPYLNKTVTGDLCPVLDMEGTGAIGAVIVLVGRDNLARCLADAEFQW
ARRSRLEAPAPGVSPPEYPGDPAHGLMTPVGMLEFEOGALLGGRSFFSLDRHPWT
PAEGDP"
4856..4861
complement(4879..4884)
polya_signal
polya_signal
complement(8315..8320)
9469..10167
/gene="UL1"
9469..10167
/gene="UL1"
/note="complexes with gH; involved in membrane fusion"
/codon_start=1
/product="viral glycoprotein L"
/protein_id="AAU88066.1"
/db_xref="GI:52843275"
/translation="MAVLGCVLLASLIFRCSPNLLSTPADSSTTEYLIGSAAATV
ADLVKGCIDLPADGVRYEARPSIDYARLITFLRYHQGCDTILMDTQABAAV
NPLFTTAGFLIEDSHALFPDNLATETARLTYKEVRLAAVASRGAASDTTVPGCYSS
EYSTRKCLGRDTPGLFDEPRLRSLAPGDEAAPPFSPAAPTSSAPPARRDRKPRP
KPRNSTEPPRRP"
10235..10990
/gene="UL2"
10235..10990
/gene="UL2"
CDS

polya_signal
polya_signal
gene
CDS
/gene="UL2"
/note="proposed start codon is GTC"
/codon_start=1
/transl_except=(pos:10235..10237,aa:Met)
/product="viral-DNA glycosylase"
/protein_id="AAU88067.1"
/db_xref="GI:52843276"
/translation="MTFDSPARPALGDDTTPRAGPPSLDMPERRRFLVDAMPPLLE
PELANPLTALAAETERRCAEVLPPREDSWITCTPDEVRVITIGDPYHQRQ
AGHLASVRRGTAPPSLNITLAVNCCBAATMSGCLLEKARGVLLNTLTAVR
RGPAPHAQIGMDRFVGSVVRLLAASRPGILVFMGMAHQAOSAIRPDPVIRVLTAVP
SPLSRVFCSCRFCLANQYLRRSLAPIDWS7"
11076..11081
11130..11813
/gene="UL3"
11130..11813
/gene="UL3"
/note="colocalizes with ICP22 and UL4 in small, dense
nuclear bodies"
/codon_start=1
/product="UL3"
/protein_id="AAU88068.1"
/db_xref="GI:52843277"
/translation="MSAVGAGAVPTLAVASGNAFAPHPDPPAPARAPAPAE
PTGHPVPPRADDRAAPGRVAPDDHYAPPTMVSVDDELGRQLDITRKDLRLSV
NLAIATKTSRSGTGKARPRRRRGRHREPPSHSLQMFVLICQPDATRVDDQRA
VIASRPRKRYTRSSDGRISPAPVVFHFVSDAPVHRDNLVAVVPPRRSTBLA
DPASDPPWQ"
11809..11814
complement(11832..11837)
polya_signal
polya_signal
gene
CDS
/gene="UL4"
complement(11897..12511)
/gene="UL4"
/note="colocalizes with ICP22 and UL3 in small, dense
nuclear bodies"
/codon_start=1
/product="UL4"
/protein_id="AAU88069.1"
/db_xref="GI:52843278"
/translation="MSASASATIAVSLGHAATTSACVLPDAEQVCAEFSGTALAS
RGLRLDALRGAVVVVRPGILVWDCRTECAVFGSRGRORLEHMDTLCAV
PPSSVSRSGSASRPTGACIATVVGDSISYITLVVGSPPCEGAPRRLPPRPD
APDAPSPVAVSPASATGAADLLEVMKEIQLSPTLVGFPDCA"
complement(12536..15163)
/gene="UL5"
complement(12536..15163)
/gene="UL5"
/codon_start=1
/product="component of helicase-primase complex"
/protein_id="AAU88070.1"
/db_xref="GI:52843279"
/translation="MTTGESAGDASAPPEPBPVPSARFDEAFNFTSMGIOP
IRVRLAGHRLLEPARVPRPDRVDALETPADLPARPPAVVLTITNAGSGKST
VTLANVEDCVVTGARITAAQNNYKLSAFLSRPINTI FHEGFGNHYQADGHP
YALASPAVEDLQERDLTYWEVADITRRVAAAGGTEDGGEFRALAALEKTLG
ABGSLTRLAFTHGALPAPTRNSIVITDAAGLGRLLTAAVYCMWMTALAEKTYA
AGLRPVLVCGSPQTQASLESTFEGKLSRQSENVITYLICNTLLEBYALPRSM
AIFINNKRCVHEFGNLMKALEGDLTEHQFVDRFVPSFETNPANLPQWTRLF
SSKHEVSAVMARLHAHLKVAREGEFVFTLPVLITAGAPPEVRELTHOPLATVEKM
IIMAGRITINYQSQDQDGRVRFVTHSQQLVAVANDTYVLNSGIATTTLRKLVF
GFGGTRSFEEAVLRDQSFVTOGBMLVERAYRFLSLITGGLIHFYFLRGLDLAK
RAAYERMALTTLLALGAEAGAGGAAPAGGRADFTLGAARAAGDAEPEDD
LNVFAGLSEOEIDLFCYHVEQDPTTAVAHQFGLKRAFLGRYHILRELFGAFAE
GAPFTYVNDNTFEGCEVWIGMRGLMSVALQOTDSYTLTGFTYAVVAFAEVRERH
AAAGLAEILEESPFPVVLVDRDHGPMASVNTVSEVESVEABELAMATNAYGTSMS
LMTITRSGSLDLKVAICTPTGNLRANSAYVAMSTTSSSEFLRANLNRERHEDD
VISGHTLSLRPELVITY"
15162..17198
/gene="UL6"
15162..17198
/gene="UL6"
/note="role in DNA cleavage/packaging"


```
/codon_start=1
/product="capsid protein"
/protein_id="AU88071.1"
/db_xref="GI:52843280"
/translation="MAERHPRGLRPPAMPDPSGPPADGCGVQHPPTMTLPRALIS
GELGTYEGGVYVNVSSSEATROLOAAVVGALLNATRHDLDEADHNRHVAARLEPE
RLVRVGGADADVAVAEVEFDTWKGLASLLDPARGLADFACCGPGSGVFSRY
IDWLKGLVPMVRPREGEVTRQLEAFILGHALPRDLATVGGAVEAAPTLALARA
FDSARIADYRVCLEFYHPRGEWRVRDPAGRGEGCVLMPPLMAGDRLLFDSMORL
SHEIVACRALREHARVCRVRNAASVKVLGRKNDGPGAPGAARASRALGEDDASRA
GSAASRLVRLIIMKGMRIHIGDINDTVRAITLDEAGHIMDSATVDALPGRAAGAC
BAAGADPRAPRROOQROARFQTSVANNINMLEGYINNLFTIERLERTYTDATQLR
DRDLERARARGLDLDRQADJGRBPGGAAPDDRADYDIIIVSKAMDVTYVAN
SFQHYVPSYADLERLSRLMEHLRCFKISHTNNGRETAIFVSGAIALFVAPY
FADVLAPITGALITGPSGLGEELEMAIFKTRVQTYLTDLALFVADVRRAARR
LPHEAADPGRSRSPVDAGRPGARPEDAVGSDHGRPELRRGGDGRP"
17149..18039
/gene="UL7"
/contig="UL7"
/length=17149..18039
/feature="UL7"
/feature="role in DNA cleavage/packaging"
/codon_start=1
/product="capsid protein"
/protein_id="AU88072.1"
/db_xref="GI:52843281"
/translation="MAAPSSGEBETAAALROAVADRDLVGAEAISHQALRMACE
VROAGRDVAFRTAGSVDPDRLRFDLGGPADLADSDYFARCGQAHKGF
AFVAVLAGEHDVSLCVPPLVPHRLSVPRPELTLELACLIMFVENCPRATHSA
FKVISAMIAVAGBRTSPDRVRLILRSQOMLNTLMFVHYVPPDRDRLPHMKAR
YLLDHPPLITLALFRAABSRVRLVRPETNLECAVYNDGVLSGMAEDFRAGLA
AMWASDAFKQTALFIRFY"
18062..18067
complement(18062..18210)
complement(18213..20489)
/gene="UL8"
complement(18213..20489)
/gene="UL8"
/codon_start=1
/product="component of helicase-primase complex"
/protein_id="AU88073.1"
/db_xref="GI:52843282"
/translation="MGBDGLAMABESVCATVLYAALPPTTRRLIALYLYLCRDARG
DAVARPAEVAVASSELQAFYGPGEISAGVAAARAASAPALPTLNDPALMRALY
ASALALEREVGAPLFVPARVGRDPRGLVARIEAPRGVPARGALGVADQVAT
DPLALAHVAHAPARALAMARLAIODTPQASASIALVRIIDTGRFAREYTLTFP
PVKEGAVGVDFEVEAVLRRPGHAQAVTARVIVPGEYFAVRAEGSPALVAFR
OMHLYHAARALPPIFALGPEPEPRGCAVDFPILGPGPPTLYVPAGAAVPAAR
LTRAHLAGAMPAPAGARALAPARILMAAAAGARGRLSPALRRPATTGLRDLDPAYTG
PVCLEARFEPGLEARLAAALCGLEPRADPRDPLARGACWMLADARPMARAVAA
LALAEAREBCPAFAVLDGVAAVGLREMAADVDLAVCGNGASILMGVLPDPCA
AGPAVERARAAALEVAGELAGGLRPEEPSPLELGIYTHAVLMGOTGAMPYNA
RTDADLEGEPORGPAYVAAAAATRRALGAAVAAVGAEGAEARALAEALDGLVLE
AFARLDAEYMAVAAEGSPDGPDPAPAAALRGALLSEHSRRVAVVCGGPDGSAV
PIDLYPRPIVLPIDCAHHRDVLAEAEVLRGALLMGEDGFAYPPEBRPLFLPD
PSGAAGAAHVAAGSGSKA"
20544..23183
/gene="UL9"
Query Match 13.3%; Score 128.4; DB 14; Length 150715;
Best Local Similarity 47.5%; Pred. No. 0.012;
Matches 381; Conservative 0; Mismatches 421; Indels 0; Gaps 0;
```

Db 125957 CGAGCCCGTGTGCGGCGCGCACGCTGCTGCGATTCGCGCGG 65

polyl₉_signal
polyl₉_signal
gene
CDS

```
QY 186 CGGAGACGGGACCTGCTCTGCTGCGCGGTACACGAGACGTTGACCGCTGAC 245
Db 125837 CGTCTCGGGCCCGGCTCTTGCGCGCGCTTCAGAGCGCGCTGCGCGCTTACCCCA 125778
QY 246 GCCCGTACCCCGGCTCATCGCGCGCGCGCGCTGCTGCGGACACCTTCTCGTGAC 305
Db 125777 CGGCGCGCGCGCGCTCTGCGCGCGCGCGCGCAACGTCGCTACCGCGTGCACGCGCT 125718
QY 306 GGGCATAGCGCGGGAACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 365
Db 125717 GGGCGCGGACACCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125658
QY 366 GATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 425
Db 125657 GCTCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125598
QY 426 CGGCGCGGCGTCACGCGCGCTGCTGCGGCTGCTGCGAGGCGCGCGCGCGCGCGCGTAC 485
Db 125597 CTTCCCGCGCGCGGAGCGCACCTGCGACCGCGCGCTGCGCGCGCGCGCGCGCGCG 125538
QY 486 GCTCAGCGGAGAGAGACGACGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 545
Db 125537 GCTGCGCGCGCTTACCTGCGCGCTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125478
QY 546 GCTCTCTCTCTTCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 605
Db 125477 GCGCGCGCGCGCTGCGCGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125418
QY 606 GACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
Db 125417 GCGCGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125358
QY 666 CCGCGAGGTACGCGGACATCAGCGGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 725
Db 125357 CTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125298
QY 726 GCTCGCGGAGCGCGCTGCTGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 785
Db 125297 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125238
QY 786 CGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 807
Db 125237 GAGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 125216
```

RESULT 12
AC149968
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 136753)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alshrocks,S.L., Amarantunge,H.C., Are,J.R., Ayele,M., Banke,T.,
Barbora,J., Benton,J., Binage,K., Blankenburg,K., Bonin,D.,
Bouck,J., Bowe,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,
Bunney,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Davila,M.L., Davis,C., Davy-Carroll,I., Dederich,D.A.,
Delaney,K.R., Delgado,O., Denna,A.L., Ding,Y., Dinh,H.H.,
Douthwaite,K.J., Drepper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,
Falls,T., Ferraguto,D., Flaggs,N., Ford,J., Foster,P., Frantz,P.,

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

2. (bases 1 to 136753)
 Worley, K.C.
 Direct Submission
 Submitted (25-JUN-2004) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 ----- Genome Center

Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information -----
 Center project name: SPCA
 Center clone name: R3-32A21
 ----- Summary Statistics -----

Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 188020 bases at least 40
Consensus quality: 159713 bases at least 30
Consensus quality: 130554 bases at least 20
Estimated insert size: 129513; sum-of-configs estimation
Quality coverage: 5x in 920 bases; sum-of-configs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hbrc.bcm.tmc.edu/docs/genbank_draft_data.html)
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of 'N', but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

*	1	1101:	contig of 1101 bp in length
*	1102	1201:	gap of unknown length
*	1202	2619:	contig of 1418 bp in length
*	2520	2719:	gap of unknown length
*	2720	4080:	contig of 1361 bp in length
*	4081	4180:	gap of unknown length
*	4181	5491:	contig of 1311 bp in length
*	5492	5591:	gap of unknown length
*	5592	6724:	contig of 1133 bp in length
*	6725	6824:	gap of unknown length
*	6825	8092:	contig of 1268 bp in length
*	8093	8192:	gap of unknown length
*	8193	9564:	contig of 1372 bp in length
*	9565	9664:	gap of unknown length
*	9665	10777:	contig of 1113 bp in length
*	10778	10877:	gap of unknown length
*	10878	14450:	contig of 3573 bp in length
*	14451	14550:	gap of unknown length
*	1451	18887:	contig of 4337 bp in length
*	18888	18987:	gap of unknown length
*	18988	20660:	contig of 1673 bp in length
*	20661	20760:	gap of unknown length
*	20761	31112:	contig of 10352 bp in length
*	31113	31212:	gap of unknown length
*	31213	47007:	contig of 15795 bp in length
*	47008	47107:	gap of unknown length
*	47108	64347:	contig of 17240 bp in length
*	64348	64447:	gap of unknown length
*	64448	85077:	contig of 20630 bp in length
*	85078	85177:	gap of unknown length
*	85178	136753:	contig of 51576 bp in length

```

FEATURES
source
location/Qualifiers
1..136753
/organism="Strongylocentrotus purpuratus"
/mol_type="genomic DNA"
/db_xref="taxon:7668"
/clone="R3-32a21"

```

ORIGIN

Query Match 12.6%; Score 122.4; DB 2; Length 136753;
 Best Local Similarity 46.4%; Pred. No. 0.037;
 Matches 427; Conservative 0; Mismatches 481; Indels 12; Gaps 4;

QY 8 GCTTCCCCCAGCGTGTGTGTGCGCGGAGCGGCGGTGGCGGATGTCGCGGCG 67
 DB 3858 GCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 37199

QY 68 TGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 127
 DB 3798 CCGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 37319

QY 128 CGGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 187
 DB 3738 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 36719

QY 188 GGGAGCGGAGCT---GTCCTGTGCGCGGTACGAGAGCGTGGAGCTTGAAGCGCTGG 244
 DB 3678 CCGGCGGAGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 36119

QY 245 ---CGCCCGTGAACCGGCTCATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 301
 DB 3618 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 35519

QY 302 GGAAGGAGATGAGCGCGGAGGAGCTCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCTCA 361
 DB 3558 GCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 34919

QY 362 ACCCGATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG---CGTGGTC 417
 DB 3498 CCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 34319

QY 418 ACCAGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 477
 DB 3438 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 33719

QY 478 CCGGATGAGGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 537
 DB 3378 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 33119

QY 538 CAGCGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 597
 DB 3318 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 32519

QY 598 --CTGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 655
 DB 3258 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 31919

QY 656 TGGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 715
 DB 3198 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 31319

QY 716 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 775
 DB 3138 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30719

QY 776 ACCGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 835
 DB 3078 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 30119

QY 836 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 895
 DB 3018 NNNCCNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 89519

QY 896 CCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 915
 DB 2958 CGTCCGCGCTCTCTGCGCTG 29319

RESULT 14
 LOCUS PM11h12g/c 1393 bp DNA linear STS 29-MAY-2003
 DEFINITION Penicillium marneffei STS, clone pm11h12.g, sequence tagged site.

ACCESSION AL684264
 VERSION AL684264.1 GI:19338009
 KEYWORDS STS;
 SOURCE Penicillium marneffei
 ORGANISM Penicillium marneffei
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Penicillium.

REFERENCE 1
 Yuen, K.Y., Pascal, G., Wong, S.S., Glaeser, P., Woo, P.C., Kunec, F.,
 Cai, J.J., Cheung, B.Y., Medigue, C. and Danchin, A.
 Exploring the Penicillium marneffei genome
 Arch. Microbiol. 179 (5), 339-353 (2003)

REFERENCE 2
 PubMed 12640520
 2 (bases 1 to 1393)
 Danchin, A. and Pascal, G.
 Direct Submission
 Submitted (08-MAR-2002) Danchin A., HKU-Pasteur Research Centre,
 Dexter HC Man Building 8, Sassoon Road, Pokfulam, Hong Kong

FEATURES
 source
 1..1393
 /organism="Penicillium marneffei"
 /mol_type="genomic DNA"
 /db_xref="taxon:37727"
 /clone="pm11h12.g"

ORIGIN

Query Match 12.3%; Score 119.6; DB 1; Length 1393;
 Best Local Similarity 45.6%; Pred. No. 0.45;
 Matches 419; Conservative 0; Mismatches 485; Indels 14; Gaps 4;

QY 60 CCGCGGCGGCTGTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 119
 DB 1064 CCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1005

QY 120 GGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 179
 DB 1004 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 945

QY 180 CCGCTTCGAGAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 239
 DB 944 GCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 885

QY 240 CGTGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 299
 DB 884 GCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 825

QY 300 CCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354
 DB 824 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 765

QY 355 GGCCTTCAACCGGATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 414
 DB 764 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 705

QY 415 GTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 474
 DB 704 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 645

QY 475 AGGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 533
 DB 644 GCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 585

QY 534 GAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 593
 DB 584 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 525

QY 594 GGCCTTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 653
 DB 524 CCGCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 465

QY 654 GCTCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 710


```

gene      complement (12172..14820)
          /gene="UL5"
          /complement (12172..14820)
CDS       /gene="UL5"
          /codon_start=1
          /product="component of helicase-primase complex"
          /protein_id="AAP41423.1"
          /db_xref="GI:30844247"
          /translation="MARVGGDASVSSPPDPREPPOFREDATINFTSMHGKSTQPTL
          VRLAERLEAAVPRLOMFRDVALEFTRPGLTREPFPVITITGNAGGKSTQPTL
          LNEVLCVVTGATRIAAQNNYVKLSGAFSLRPINT1PHEFGFNNVQAVQGHPTTL
          ASNPASVEDLQRRDLTYMEVIADITRAVAAGGAGDEGNEFRALATLESTGLEG
          ALTRLAFTGALPAFTRSNVIIDEGKLGRLHLAAVYMMMLATYTPPYAAGL
          RPLVCGSPSTOTASLESTFEHOKLRCSIRSENVLTYIICNLTREARLPRSMALF
          INKRCVHEHFRGMLKALEYGLPTEBHOFPVRPVPVPSFPTINPNALPQMTLPSGH
          KEYSAYMARLHAHLKVARBGEFVFSLPVITVSAGAFDEYRELTHOPALTVKMLAA
          NAGRITNYSQSDQDAGRARLEVHSKQQLVANRDVTVLNSQ1AATTRLRKLPFES
          GTRPSEAVLRDSDPVKTOGEMLVERFAYOFLSRILFGGLHFPYFLIRPGLDARAR
          AYERMGALTRIELSLGATCGAPATPGQOASGGESASAPGARAFDFKTLGARATD
          AEPEDDLVVPFAGLNEOEIDLPHYCHYTVOPEPTTAIVHAQFLKRALFLGRYHILRE
          LFEAGFEGAPGTYVNDNTPRGCENMISGRGLMVALOTDSYTLNGPTXAVYAFAP
          BEYRRRAAGGLALEBESPLPYVLRDQGFMSVYNTNISFVESVESRELMATINA
          DVGISMLANTITRSQGLSDKVAICTPQGNLRNLNSAYVAMSRTTSSEFLRNMLNPLR
          BRHERDVISEHLSALRDPDLIVY"
          14819..16870
          /gene="UL6"
          /db_xref="GI:30844248"
          /protein_id="AAP41424.1"
          /db_xref="GI:30844248"
          /note="role in DNA cleavage/packaging"
          /codon_start=1
          /product="capsid protein"
          /protein_id="AAP41425.1"
          /db_xref="GI:30844249"
          /translation="MATSTPAEBETAAALRRVAGDRDLVGVAEIASHOALLRNACE
          VROATHHTVFTATGVAVNVVTPDRLHLFVLDCGPDALIASDAVFAKRDOSARHGF
          ARAVLARSDHSHSLCVPPLVLEPHRLSVFPEHLIEDFELACLIMYENOPRAATASA
          FVKISAMLAALGRRTSPFDRVLCILRS COMMLINTLMFMVHVEPFDQVLPKMCAR
          YLLANNPPLITLTPRAGPSRVFRLFPQPRASLSECAVYNPQGLAAGMASDEPRAGLV
          SWSSDAPKTOGSLFYLIV"
          complement (17886..20171)
          /gene="UL8"
          /complement (17886..20171)
          /gene="UL8"
          /codon_start=1
          /product="component of helicase-primase complex"
          /protein_id="AAP41426.1"
          /db_xref="GI:30844250"
          /translation="MGDAGLVMAESICAITTYTAWLPRTDRDLVLLYLPFCRDARG
          DGAARPAEVSIASTELQAFYPAEISAGAVAAARAASAPAAPLETTLNDPVLKMRALY
          ACALVALEREVGRALFAPARVAMDPRGTGIEGLPGORSAPPAALDVNAQIAT
          DPLALAAHVAAARPDARLAWARLAAIDQTQCSAASALTVIDGAAARFARERTTLTFP

```

```

PVRRBGAADVFEVREAVLRPRGHAQVATVAVLPVPGYEFVAVGAFSAPALIAMER
OHVLTVAHAGRALPPIFAFIAPFEBRGAQVNDHESVIFGRGMATLTKVSGAAAPPA
LAAHAEIAGAMPAAAGARALGPRARMAAVSAAABERLPALRRPRAAGLILDPAYIG
PVLARFRFGLQPLALVLEIGLGRADPRADPRDLGRGARMRDLGEBGAEOA
PMPREVLAATDEVRREYPAVARITLDGIVSAVGLKKAAGVADVAACDGGGRVGV
LNVDPDAAGAGAAAAAERARALAAASSEALAGVGLDAPPLVLGRTYTHAVLVM

Query Match      12.1%; Score 117.4; DB 14; Length 156789;
Best Local Similarity 46.2%; Pred. No. 0.084;
Matches 391; Conservative 0; Mismatches 456; Indels 0; Gaps 0;

Qy      6  CGGCTTCCCCCGGAGCGTGTGTCGCGCGGAGCGGGGGGCGGATGTTCCGGG 65
Db      153893  CGCGCGCGCGCGCGCGCGCTGTGCTGTGACCGCGGCGCCCGAGAGAGTGGCGGCCA 153952

Qy      66  GCTGTCCGAGAGCGGAGACCGGACGCTGTGTGACCTTGTCACCGCGCGGAGCG 125
Db      153953  CGGGCCCGCGTGTGCGCGCGCGCCACGTTACCTGCGCTGCGCGCGCGCGCGCA 154012

Qy      126  GCCGAGCGCTGCTGTGTGGGAGACTCACCGCGCGGGGGCCCGAATCGCGCGCGCTT 185
Db      154013  GTGCGACGTGCGTGTGCGCGCGCGCGCGGACCTGCGCGCGCGCGCTGTGCGCGCG 154072

Qy      186  CGGAGACGCGGACCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 245
Db      154073  CGTCTTCCGGGCGCGCGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154132

Qy      246  GCCCGTACCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 305
Db      154133  CGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154192

Qy      306  GGGCATGTGCGCGGAGCTGCGCGGCCCAAGCCCGCGCGCTGCTGCTGCTGCTGCTGCT 365
Db      154193  CGGGCCCGACACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 154252

Qy      366  GATGTTGCGCGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 425
Db      154253  GCTGACGCGCGCGCGCGCGCGGATGATGATGATGATGATGATGATGATGATGATGAT 154312

Qy      426  CGGGCGGGGCGTCACGGCGCTGTGCGAGCTGTGCGAGCGGCGGCGGAGCGCGCTAGC 485
Db      154313  CTTCCGCGCGCGGAGGCGGACGACGCGGCGGCTGCGCGCGCGCGCGCGCGCGCGCG 154372

Qy      486  GCTGACGCGGAGAGACGACGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 545
Db      154373  GCTGCGGCGCGCTCTACTCTGCGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 154432

Qy      546  GCTCCTCTCTCTGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 605
Db      154433  GCGCGCGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 154492

Qy      606  GACGCGACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 665
Db      154493  GCGCGCTCGTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 154552

Qy      666  CCGCGAGGTGACGCGGAGCATTCAGCGGTCAACCCCGGCGGCGGTCCGCGCGCGCGCG 725
Db      154553  GCGCGGCGGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 154612

Qy      726  GCTCGCGGAGGCGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 785
Db      154613  GCGCGCGCGCGGAGCGCGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 154672

Qy      786  CGCGCGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 845
Db      154673  CGCGCGCGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 154732

Qy      846  CGACGCGC 852
Db      154733  CGCGCGC 154739

Search completed: October 5, 2005, 05:30:57

```

Job time : 4458.25 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 16:46:09 ; Search time 585.401 Seconds
(without alignments)
9798.799 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969

Sequence: 1 atgcgcgcgtcccccgcag.....aagcgagaaagcagatga 969

Scoring table: IDENTITY_NUC

Searched: 4390206 seqs, 2959870667 residues

Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

N_Geneseq_16Dec04:*
1: geneseqn1980s:*
2: geneseqn1980s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001as:*
6: geneseqn2002as:*
7: geneseqn2002as:*
8: geneseqn2003as:*
9: geneseqn2003as:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES.

Result No.	Score	Query Match	Length	DB	ID	Description
1	969	100.0	969	4	AAF86411	Aaf86411 4-amino-4
2	969	100.0	969	8	AAL50182	Aal50182 S venezue
3	915.4	94.5	3305	8	ABZ69799	Abz69799 Plasmid p
4	913.8	94.3	12391	8	ABZ69798	Abz69798 Plasmid p
5	270.8	27.9	888	2	AAT59269	Aat59269 Streptomy
6	270.8	27.9	2888	2	AAT59268	Aat59268 Streptomy
7	124.8	12.9	114955	4	AAD17166	Aad17166 Human ade
8	115.6	11.9	125401	4	AAD17166	Aad17166 Streptomy
9	115.4	11.9	114955	2	AAK53491	Aak53491 Human ade
10	114.8	11.8	65140	4	AAD17184	Aad17184 Streptomy
11	111.8	11.5	9222	8	AAL61171	Aal61171 Actinosyn
12	111.6	11.5	82746	8	AAL61224	Aal61224 Actinosyn
13	111.6	11.5	4770	8	AAL61172	Aal61172 Actinosyn
14	109.8	11.3	9975	8	AAK54645	Aak54645 Streptomy
15	108.8	11.2	113193	8	AAK54645	Aak54645 Streptomy
16	107.4	11.1	969	4	AAF86411	Aaf86411 4-amino-4
17	107.4	11.1	969	4	AAF86411	Aaf86411 4-amino-4
18	106.8	11.0	14055	8	AAL50182	Aal50182 S venezue
19	106.2	11.0	3305	8	ABZ69799	Abz69799 Plasmid p
20	105.4	10.9	12391	8	ABZ69798	Abz69798 Plasmid p

21	105.2	10.9	3354	8	AAL61181	Aal61181 Actinosyn
22	104.4	10.8	44377	2	AAT80414	Aat80414 Platenoli
23	104.4	10.8	44377	2	AAT80414	Aat80414 Platenoli
24	104.4	10.7	13842	3	AAZ87297	Aaz87297 S. venezu
25	104	10.7	13842	12	ADL91915	Adl91915 Streptomy
26	104	10.7	36778	3	AAZ87318	Aaz87318 S. venezu
27	104	10.7	36778	12	ADL91933	Adl91933 Streptomy
28	104	10.7	37948	3	AAZ87285	Aaz87285 S. venezu
29	104	10.7	38506	3	AAZ87285	Aaz87285 S. venezu
30	104	10.7	38506	3	AAZ87285	Aaz87285 S. venezu
31	104	10.7	38506	8	ADA09418	Ada09418 Cosmid pk
32	104	10.7	38506	10	ADH53462	Adh53462 S. venezu
33	104	10.7	38506	10	ABSS6090	Abss6090 S. venezu
34	102.6	10.6	3957	5	AAK09686	Aak09686 HSV-2 imm
35	102.6	10.6	3957	10	ADG75121	Adg75121 Human her
36	102.6	10.6	154746	6	AAD25519	Aad25519 Human her
37	102.6	10.6	154746	6	AAD25519	Aad25519 Human her
38	101.4	10.5	5802	6	ABZ86696	Abz86696 S. kaniba
39	101.4	10.5	15738	10	AD123932	Ad123932 Streptomy
40	101.4	10.5	61944	10	AD123920	Ad123920 Streptomy
41	101.4	10.5	109519	5	AAK08693	Aak08693 Micromono
42	100.4	10.4	2700	4	AAD17193	Aad17193 ERD4 ins
43	100.4	10.4	110000	4	AAL19682	Continuation (40 o
44	100.2	10.3	58857	3	AAK58471	Aak58471 Nucleotid
45	100	10.3	88421	6	AAL40781	Aal40781 Nucleotid

ALIGNMENTS

RESULT 1	AAF86411	standard, DNA, 969 BP.
ID	AAF86411	
XX	AAF86411	
AC	22-JUN-2001	(first entry)
DT		
XX		
DE	4-amino-4-deoxyphenic acid dehydrogenase coding sequence.	
XX		
KM	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;	
KW	4-amino-4-deoxyphenic acid dehydrogenase; enzyme; pnpC; ds.	
XX		
OS	Streptomyces venezuelae.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..969
FT		/product= "4-amino-4-deoxyphenic acid dehydrogenase"
XX		
PN	W0200123542-A1.	
XX		
PD	05-APR-2001.	
XX		
PF	29-SEP-2000; 2000MO-JP006783.	
PR	XX	
XX	29-SEP-1999; 99JP-00276314.	
XX		
PA	(MEIJ) MEIJ SEIKA KAISHA LTD.	
XX		
PI	Yanaí K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Midoh N;	
PI	Murakami T;	
XX		
DR	WPI; 2001-290517/30.	
DR	P-PDB; AAB82073.	
XX		
PT	Transformant producing secondary metabolite modified with functional	
PT	group e.g. benzene with nitrogen-containing substituent at para-position,	
PT	FP1022, with ease at low cost, for application in pharmaceuticals and	
XX	agrochemicals.	
XX		
PS	Claim 15, Page 67-70; 83pp; Japanese.	
XX		

Best Local Similarity 100.0%; Pred. No. 2.8e-113;
Matches 969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATAGAGGAGCTTCCCGGAGAGCTGCTGTCGAGCGGAGCGGAGCGGATGTC 60
DB 1 ATAGAGGAGCTTCCCGGAGAGCTGCTGTCGAGCGGAGCGGAGCGGATGTC 60
QY 61 GCGGAGCTGCTGCGGAGAGCGGAGCGGAGCGGAGCTGCTGTCGAGCGGAGCGG 120
DB 61 GCGGAGCTGCTGCGGAGAGCGGAGCGGAGCGGAGCTGCTGTCGAGCGGAGCGG 120
QY 121 GAGCGGCGGAGCGGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 180
DB 121 GAGCGGCGGAGCGGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 180
QY 181 GCGCTCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 240
DB 181 GCGCTCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 240
QY 241 GTCGCGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 300
DB 241 GTCGCGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 300
QY 301 CGGAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 360
DB 301 CGGAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 360
QY 361 AACCCGATGTTCCCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 420
DB 361 AACCCGATGTTCCCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 420
QY 421 AGGAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 480
DB 421 AGGAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 480
QY 481 GTACGCTCAAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 540
DB 481 GTACGCTCAAGCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 540
QY 541 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 600
DB 541 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 600
QY 601 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 660
DB 601 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 660
QY 661 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 720
DB 661 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 720
QY 721 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 780
DB 721 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 780
QY 781 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 840
DB 781 GCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGG 840
QY 841 AACCTCGAGCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGG 900
DB 841 AACCTCGAGCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGG 900
QY 901 CAGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAG 960
DB 901 CAGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAGCTGCTGTCGAGCGGAG 960
QY 961 GACCGATGA 969
DB 961 GACCGATGA 969
```

RESULT 3

```
ABZ69799
ID ABZ69799 standard; DNA; 3305 BP.
XX
AC ABZ69799;
XX
DT 08-APR-2003 (first entry)
XX
DE Plasmid papABC.
XX
KW DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
KW biosynthesis; p-aminophenylalanine; pAF; chorismate.
XX
OS Synthetic.
XX
PN WO200285923-A2.
XX
PD 31-OCT-2002.
XX
PF 19-APR-2002; 2002WO-US012465.
XX
PR 19-APR-2001; 2001US-0285030P.
XX
PR 06-FEB-2002; 2002US-0355514P.
XX
PA (SCRI ) SCRIPPS RES INST.
XX
PI Schultz P, Wang L, Anderson JC, Chin JK, Liu DR, Magliery TJ;
PI Meggers EL, Mehl RA, Pasternak M, Santoro SW, Zhang Z;
XX
DR WPI; 2003-120430/11.
XX
PT Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
PS Example 4; Page 127-128; 188pp; English.
XX
CC The invention relates to a novel composition comprising a translation
CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC -tRNA with at least one unnatural amino acid in the translation system
CC and the O-tRNA recognises at least one selector codon. A composition of
CC the invention is useful for producing at least one protein comprising at
CC least one unnatural amino acid. The protein is the Aspi2TAG mutant of
CC chloramphenicol acetyltransferase (CAT); the Tyr163TAG mutant of mouse
CC dihydrofolate reductase (DHFR), or a Tyr163TAG mutant of mouse
CC dihydrofolate reductase comprising CHO-His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid containing the individual genes papABC that encode
CC the enzymes used to carry out the conversion of chorismate to the
CC unnatural amino acid p-aminophenylalanine (pAF)
XX
SQ Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;
Query Match 94.5%; Score 915.4; DB 8; Length 3305;
Best Local Similarity 96.6%; Pred. No. 1.1e-106;
Matches 953; Conservative 0; Mismatches 16; Indels 18; Gaps 1;
QY 1 ATAGAGGAGCTTCCCGGAGAGCTGCTGTCGAGCGGAGCGGAGCGGATGTC 60
DB 2007 ATAGAGGAGCTTCCCGGAGAGCTGCTGTCGAGCGGAGCGGAGCGGATGTC 2066
QY 61 GCGGAGCTGCTGCGGAGAGCGGAGCGGAGCGGAGCTGCTGTCGAGCGGAGCGG 120
DB 2067 GCGGAGCTGCTGCGGAGAGCGGAGCGGAGCGGAGCTGCTGTCGAGCGGAGCGG 2126
QY 121 GAGCGGCGGAGCGGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 180
DB 2127 GAGCGGCGGAGCGGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2186
QY 181 GCGCTCGGAGAGCGGAGCTGCTGTCGAGCGGAGCGGAGCGGAGCGGAGCGG 240
```

```
Db 2187 GGCCTCCGGAGACCGGACCTGCTCTGCTCCGCTACACGAGGAGCGGCGCTCAAGGCC 2246
Qy 241 GTGGCGCCCGGTGACCCGGGCTCATGGGCGGGGCGCGCTGCTGCGCGACACCTGTCGCTC 300
Db 2247 GTGGCGCCCGGTGACCCGGGCTCATGGGCGGGGCGCGCTGCTGCGCGACACCTGTCGCTC 2306
Qy 301 CGGACGGGAGATGGCGCGGAGCTCGGCGGCCACGCCCGCGCTGCGAGCATGGGCGCTC 360
Db 2307 CGGACGGGAGATGGCGCGGAGCTCGGCGGCCACGCCCGCGCTGCGAGCATGGGCGCTC 2366
Qy 361 AACCCGATGTTGCCCCCGCGCGGATGACCGGCGGACCCGTCGCGCGCTGTCAC 420
Db 2367 AACCCGATGTTGCCCCCGCGCGGATGACCGGCGGCGCGCTGTCAC 2426
Qy 421 AGGGAAGGCGCGGCGCTGACCGGCGCTGCGCGGCTGCGAGGCGCGGCGGCGGCGGCG 480
Db 2427 AGGGAAGGCGCGGCGCTGACCGGCGCTGCGCGGCTGCGAGGCGGCGGCGGCGGCGGCG 2486
Qy 481 GTACGGCTCAACGGCGGAGAGACGACCGGACGACGCGCGGCGGCGGCGGCGGCGGCG 540
Db 2487 GTACGGCTCAACGGCGGAGAGACGACCGGACGACGCGCGGCGGCGGCGGCGGCGGCG 2546
Qy 541 GCGGTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 2547 GCGGTGATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606
Qy 601 GCGGCGACGCGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660
Db 2607 GCGGCGACGCGCACCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2666
Qy 661 GCGGCGCGCGGAGGTATACGGGAGCATTCAGCGGTTCACCCCGGCGGCGGCGGCGGCG 720
Db 2667 GCGGCGCGCGGAGGTATACGGGAGCATTCAGCGGTTCACCCCGGCGGCGGCGGCGGCG 2726
Qy 721 CGGGCGCTCGCGGAGGCGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 768
Db 2727 CGGGCGCTCGCGGAGGCGCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2786
Qy 769 -----GACCGGAGCGGTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 822
Db 2787 GCGGAGAGACCGGAGCGGCGGCGGAGCGGCGGCGGAGCGGCGGAGCGGCGGCGGCGG 2846
Qy 823 GGAATGCGACGCGCGCGGGAACCTTCAGCGGCTCTTCGGGGAATCTCGCGCGGCTCAT 882
Db 2847 GGAATGCGACGCGCGCGGGAACCTTCAGCGGCTCTTCGGGGAATCTCGCGCGGCTCAT 2906
Qy 883 CGGAGCTCGCGCGGCGGCGGCGGAGCACTGCGAGAGCTGTTCCGCACTCTCCACCGCAC 942
Db 2907 CGGAGCTCGCGCGGCGGCGGCGGAGCACTGCGAGAGCTGTTCCGCACTCTCCACCG 2966
Qy 943 GACGACGAGGCGGAGGAGGAGCGATGA 969
Db 2967 GACGACGAGGCGGAGGAGGAGCGATGA 2993

RESULT 4
ABZ69798
ID ABZ69798 standard; DNA; 12391 BP.
AC ABZ69798;
XX
DT 08-APR-2003 (first entry)
XX
DE Plasmid plasc-papabc.
XX
XX DHFR; translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;
KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;
KM chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;
XX biosynthesis; p-aminophenylalanine; pAF.
XX
XX Synthetic.
```

```
PN WO200285923-A2.
XX
XX 31-OCT-2002.
PD
XX
XX 19-APR-2002; 2002WO-US012465.
PF
XX
XX 19-APR-2001; 2001US-0285030P.
PR
XX 06-FEB-2002; 2002US-0355514P.
XX
XX (SCRI ) SCRIPPS RES INST.
PA
XX Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TU;
PI Meggers EL, Mehl RA, Pastinak M, Santoro SW, Zhang Z;
XX
XX MPI; 2003-120430/11.
DR
XX
XX Composition useful for producing protein comprising unnatural amino acid,
PT has translation system comprising orthogonal tRNA and orthogonal
PT aminoacyl tRNA synthetase.
XX
XX Example 4; Page 124-127; 188pp; English.
XX
XX The invention relates to a novel composition comprising a translation
CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl
CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O
CC tRNA with at least one unnatural amino acid in the translation system
CC and the O-tRNA recognises at least one selector codon. A composition of
CC the invention is useful for producing at least one protein comprising at
CC least one unnatural amino acid. The protein is the Asp127NG mutant of
CC chloramphenicol acetyltransferase (CAT), the Tyr1637NG mutant of mouse
CC dihydrofolate reductase (DHFR), or a Tyr1637NG mutant of mouse
CC dihydrofolate reductase comprising CHOOH His6tag. The unnatural amino
CC acid is provided exogenously. The translation system is a cell and the
CC unnatural amino acid is biosynthesised by the cell. The present sequence
CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine
XX (pAF) in vivo
XX
XX Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;
SQ
XX
XX Query Match 94.3%; Score 913.8; DB 8; Length 12391;
XX Best Local Similarity 96.5%; Pred. No. 1.2e-106;
XX Matches 952; Conservative 0; Mismatches 17; Indels 18; Gaps 1;
Qy 1 ATGAGCGGCTTCCCGCGAGCGTCTGTCGCGGCGGAGCGGAGCGGTCGCGCATGTC 60
Db 2361 ATGAGTGGCTTCCCGCGAGCGTCTGTCGCGGCGGAGCGGAGCGGTCGCGCATGTC 2420
Qy 61 GCGGCGCTGTCGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 120
Db 2421 GCGGCGCTGTCGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 2480
Qy 121 GGAAGCGCGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 180
Db 2481 GGAAGCGCGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2540
Qy 181 GCGCTTCGGGAGCGCGGAGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 240
Db 2541 GCGCTTCGGGAGCGCGGAGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCT 2600
Qy 241 GTGGCGCCCGGTGACCCGGGCTCATGGGCGGGGCGGCTGCTGCGGACACCTGTCGCTC 300
Db 2601 GTGGCGCCCGGTGACCCGGGCTCATGGGCGGGGCGGCTGCTGCGGACACCTGTCGCTC 2660
Qy 301 CGGACGGGAGATGGCGCGGAGCTCGGCGGCCACGCCCGCGGCTCGAGACGTTGGGCGTC 360
Db 2661 CGGACGGGAGATGGCGCGGAGCTCGGCGGCCACGCCCGCGGCTCGAGACGTTGGGCGTC 2720
Qy 361 AACCCGATGTTGCCCCCGCGCGGATGACCGGCGGAGCGGAGCGGAGCGGAGCGGAGCG 420
Db 2721 AACCCGATGTTGCCCCCGCGCGGATGACCGGCGGAGCGGAGCGGAGCGGAGCGGAG 2780
Qy 421 AGGGAAGGCGCGGCGCTGACCGGCGCTGCTGCGGCTGCTGAGGCGGCGGCGGAGCGGCG 480
```

Db	2781	AGGAGACGGGGCCGGGGGCTACAGGACCCCTGCTCGGGCTCTGCAAGGGCGGGGGGAGAGGCC	2840
Qy	481	GTACGGCTTCAACGGCGGAGAGACGACCGGACGACGAGCGGCGGCCACCCAGAGCCCTGACGAC	540
Db	2841	GTACGGCTTCAACGGCGGAGAGACGACCGGACGACGAGCGGCGGCCACCCAGAGCCCTGACGAC	2900
Qy	541	GCCGCTCTCTTCTCTTCTTGGGGTCCGCCCTTCGCGCCGCTTCGGAGCTGACAGTCCGGGACCTG	600
Db	2901	GCCGCTCTCTTCTCTTCTTGGGGTCCGCCCTTCGCGCCGCTTCGGAGCTGACAGTCCGGGACCTG	2960
Qy	601	GCGGCGACGAGCACCGCCCGCCACAGGAGTGTCTGAGCCCTTCCTGGCCGATGTGCTCGGC	660
Db	2961	GCGGCGACGAGCACCGCCCGCCACAGGAGTGTCTGAGCCCTTCCTGGCCGATGTGCTCGGC	3020
Qy	661	GCGACGCCCCGAGGTGTACGGGGACATCCAGCGGTCCAAACCCCGGGCGAGTCCGCGGC	720
Db	3021	GCGACGCCCCGAGGTGTACGGGGACATCCAGCGGTCCAAACCCCGGGCGAGTCCGCGGC	3080
Qy	721	CGGGGCGCTTCGCGAGAGCCCTTGCGCTCTTCGCGCGCTGTGTGGTGAC-----768	780
Db	3081	CGGGGCGCTTCGCGAGAGCCCTTGCGCTCTTCGCGCGCTGTGTGGTGAC-----768	3140
Qy	769	-----GACCCGGAGACGTGTGCGAGCGCCCCCGGGCGGGCGAGACGCCCGGCGATCCGGG	822
Db	3141	GCCGAGAGACCCGGAGACCGGCGGACCGACCGACCCCGACCGACGAAACCCCGGCGATCCGGG	3200
Qy	823	GGATGCGACGAGCGCGCGCGGGAACCTTCGAGCGGCGTCTTCGGGGAACTCCGCGGCTCATGGGA	882
Db	3201	GGATGCGACGAGCGCGCGCGGGAACCTTCGAGCGGCGTCTTCGAGGAACTCCGCGGCTCATGGGA	3260
Qy	883	CCGAGCTTCGCGGCGGGGCCAGGACATTCGACGAGGCTGTTCGCAACCTTCACCGCACCC	942
Db	3261	CCGAGCTTCGCGGCGGGGCCAGGACATTCGACGAGGCTGTTCGCAACCTTCACCGCACCC	3320
Qy	943	GACGACGAAAGCGAGAACGCGATGA	969
Db	3321	GACGACGAAAGCGAGAACGCGATGA	3347
RESULT 5			
AAT59269			
ID	AAT59269	standard; cDNA; 888 BP.	
XX	AAT59269;		
AC			
XX			
DT	02-APR-1997	(first entry)	
XX			
DE	Streptomyces prietiaespiralis papC gene.		
XX			
KW	Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin; DMPABA precursor; 4-dimethylamino-L-phenylalanine; pAPa; pAPm; pAPb; pAPc; isomerisation; aromatisation; N-methyltransferase; ds.		
XX	Streptomyces prietiaespiralis.		
OS			
PH	Key	Location/Qualifiers	
XX	mat_peptide	1..888	
FT		/*tag= a	
FT		/product= "PapC"	
XX	WO9601901-A1.		
XX	25-JAN-1996.		
XX			
PF	04-JUL-1995;	95WO-FR000889.	
XX			
PR	08-JUL-1994;	94FR-0008478.	
XX			
PA	(RHON) RHONE POULENC RORER SA.		
XX			
PI	Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;		
XX	Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;		
XX			

[illegible]

Dd 29394 AGCGCTTCGCCGCGCCGTGCAGCAGCGCGACACCCGGGAACTGGCGCTGCTTGGGGGG 29455

Oy 908 ACTGCGAAGACTGTTCGCACACCTCCACCGCACCAAC 945

Dd 29454 CTTGGCCCTGACTGTGATCGGCGCCGTCCTTCGCGCGAC 29491

RESULT 9
ID AAX53491 standard; DNA; 114955 BP.
XX AAX53491;
XX AAX53491;
DT 05-JUL-1999 (first entry)
XX Human adenosine A1 receptor antisense oligonucleotide fragment.
XX Antisense oligonucleotide; multiple target; antisense treatment;
KW impaired respiration; inflammation; lung disease;
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;
KW acute asthma; allergy; asthma; impeded respiration;
KW respiratory distress syndrome; pain; cystic fibrosis;
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;
KW prostate cancer; ss.
XX Synthetic.
OS MO9913886-A1.
PN 25-MAR-1999.
PD 17-SEP-1998; 98WO-US019419.
PF 17-SEP-1997; 97US-0059160P.
XX 17-SEP-1997; 97US-0059160P.
PR 09-JUN-1998; 98US-00093972.
XX (UYEC-) UNIV EAST CAROLINA.
PA MYCE JW;
PI WPI; 1999-229400/19.
DR New antisense oligonucleotides used in treatment of, e.g. pulmonary
XX vasoconstriction.
PT Disclosure; Page 37; 120BP; English.
PS The specification describes antisense oligonucleotides (AAX52869-X55271)
XX directed against at least 2 mRNAs selected from target genes, coding and
CC non-coding regions of RNAs corresponding to target genes, gene initiation
CC codons, genomic flanking regions, intron-exon borders, the 5'-end, the 3'
CC -codon and the juxta-section between coding and non-coding regions and all
CC segments of RNAs encoding proteins associated with one or more diseases,
CC conditions or mixtures. The antisense oligonucleotides may be derived
CC from sequences AAX55272-74. These multiple target oligonucleotides
CC (specifically AAX55180-271) can be used for the antisense treatment of
CC diseases and conditions. Typical diseases and conditions are those
CC associated with impaired respiration and inflammation, including lung
CC diseases, pulmonary vasoconstriction, inflammation, allergic rhinitis,
CC acute asthma, allergies, asthma, impeded respiration, respiratory
CC distress syndrome, pain, cystic fibrosis, pulmonary hypertension,
CC pulmonary vasoconstriction, emphysema, chronic obstructive pulmonary
CC disease (COPD), and cancers such as leukemias, lymphomas, carcinomas e.g.
CC colon cancer, breast cancer, lung cancer, pancreatic cancer,
CC hepatocellular carcinoma, kidney cancer, melanoma, hepatic metastases, as
CC well as all types of cancers which may metastasize or have metastasized
CC to the lungs, including breast and prostate cancer
XX Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 0 U; 21427 Other;

```

Query Match      11.9%; Score 115.4; DB 2; Length 11495;
Best Local Similarity 34.2%; Pred. No. 6,4e-07;
Matches 321; Conservative 101; Mismatches 502; Indels 15; Gaps 3;

OY      15  CGCAGCGCTGCTGCTCGGCGCAGCGGGGCGGTGAGCGGCATGTTGCGCGGCTGCTGC 74
Db      103996  CGGCTCTCGGGCCCGGGCCSNNNNDNNNTTGGCGGCTGTCGGGCGGGGCSNNNDNNNGCGGCTCGG 104055
OY      75  GAGAGCGGAGCAGCCGACAGCTTGTGTGTGACCTGTGACCGCGCCGCGGAGCGCGGACGC 134
Db      104056  GCCGGGCSNNNDNNCGCGGCGTGGGCGCGGCGCSNNNDNNCGGCGTGGGCGCGGCGSNNDN 104115
OY      135  CTGCTCGGTGGGCGACGTCACCGCGCCGGGGGCCCGAATCTCGGGCCGCTCTCGGGACGC 194
Db      104116  NCGCGTCCGGCCGGGCSNNNDNNGCGTGGGGCCGGGCSNNNDNNCTGTGGGCGCGGCSNN 104175
OY      195  GGAAGCTGCTGCTGCTGCTGACAGAGACGTGGGCGCTCAAGAGCGGTGAGCGCCGTGAC 254
Db      104176  NDNNGTCCGGCCCGGGCSNNNDNNNTTGGGCGGGGCSNNNDNNCGGGCCGGGCSNNNDNNCC 104235
OY      255  CCGGCTCATGCGGC-----CGGCGCGCTGCTGCGCGACACCTGTCTCGTCCGAGCG 306
Db      104236  GCBGGCCBGGGCGGCGCGCGCGCGCGGCGSNNDNNCCGCBGGCCBGGGCGCGCGCG 104295
OY      307  GGCATGCGCGCGGAGCTGCGGGCCACGCGCCCGCGTCTCAGACGTGGGCTCAACCGG 366
Db      104296  GCCGGGCSNNNDNNCCGCBGGCCBGGGCGCGCCCGCGCGGCSNNNDNNCCGCBGGC 104355
OY      367  ATGTTGGCCCCCGCGCGCGGACGTACCGGCGCACCGGTGGCGCGGTGTCAACAGGAGAC 426
Db      104366  CBGGGCGGCGCGCGCCCGCGGGSNNNDNNCCGCBGGCCBGGGCGCGCGCGCGCGGCSNN 104415
OY      427  GGGCGGGGCGTCACGGC-----CCTGCTGCGGCTGTGTCGAGGCGGCGCGGAGAGCC 480
Db      104416  DNNCCGCBGGCCBGGGCGCGCGCGCGCGGCSNNNDNNCCGCBGGCCBGGGCGCGCGCG 104475
OY      481  GTAACGCTCACGCGCGAGAGACGACCGGAGCGAGCGCGGCCACCGAGGCTGTAGCGAC 540
Db      104476  GCCSNNNDNNCCGCBGGCCBGGGCGCGCGCGGCSNNNDNNCCGCBGGCCBGGGCGCGCG 104535
OY      541  GCCGTGCTCTCTCTTCTTGGGGCTGCCTCGCGCGCTCTCGGCGTGAAGTCCGGGCGCTG 600
Db      104536  CGCCGGSNNNDNNCCGCBGGCCBGGGCGCGCGCGGCSNNNDNNCCGCBGGCCBGGGCGCG 104595
OY      601  GCGGCGAGCGGACCGCGCGCCGCCACAGAGGTGTGTCTGCGCCTCTGGGCGGTGCTCGGC 660
Db      104596  CCGCCSNNNDNNCCGCBGGCCBGGGCGCGCGCGGCSNNNDNNCCGCBGGCCBGGGCGCGCG 104654
OY      661  GAGCAGCCCCGAGGTGTACGGGAGCATTCAGCGGATCAACCCCGGCGGCGTCCGCGCGC 720
Db      104655  GSNNDNNCCGCBGGCCBGGGCGCGCSNNNDNNCCGCBGGCCBGGGCGCGCGCSNNNDNNCC 104714
OY      721  CGGGCGCTGCGCGAGGCGCTGCGTCTTTCGCGCGCTGTGTGTGGGAGCGACCCGAGCGT 780
Db      104715  CGCBGGCCBGGGCGGGSNNNDNNCCGCBGGCCBGGGCGGCSNNNDNNCCGCBGGGCGC 104774
OY      781  GCCGACGCGCCCGGGCGGCGCGAGCGCCCGGCGCATTCGCGGGGAGTGCAGCGGCGCGG 840
Db      104775  GSNNDNNCCGCBGGCCBGGGCSNNNDNNCCGCBGGCCBGGGGSNNNDNNCCGCBGGCCBG 104834
OY      841  AACCTGACGCGGCTTTCGGGGAATCTCCGCGGCTCATGGAACCGGAGCTCGCGCGGCG 900
Db      104835  GSNNDNNCCGCBGGCCBGGGCSNNNDNNCCGCBGGCCBSNNNDNNCCGCBGGCCSNNNDNNCC 104894
OY      901  CAGGACCACTGCCAGAGAGCTGTTCCGACACCTCCACCGC 939
Db      104895  GCBGGCCBGGGCGCGCGCGCGCGCGGCGGCSNNNDNNCC 104933

RESULT 10
AADI1784
ID AADI1784 standard; DNA; 65140 BP.
XX

```



```
OS Actinosynnema pretiosum.
XX
XX WO2003045312-A2.
XX
XX 05-JUN-2003.
XX
XX 21-NOV-2002; 2002MO-US037547.
XX
XX 21-NOV-2001; 2001US-0332158P.
XX
XX (UNIW ) UNIV WASHINGTON.
XX
XX Flose HG, Yu T, Leistner E;
XX
XX WPI; 2003-493374/46.
XX
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
XX expression vector comprising open reading frame from ansamitocin gene
XX cluster I of Actinosynnema pretiosum, useful as antitumor agent.
XX
XX Claim 7; Page 105-152; 160pp; English.
XX
XX
XX The invention relates to maytansinoid produced by bacterial host cell
XX transformed with expression vector comprising open reading frame from
XX ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is
XX useful as a potent antitumor agent. The present sequence is A. pretiosum
XX ansamitocin ansamitocin biosynthetic gene cluster I
XX
XX Sequence 82746 BP; 10231 A; 31648 C; 32359 G; 8508 T; 0 U; 0 Other;
SQ
Query Match 11.5%; Score 111.8; DB 8; Length 82746;
Best Local Similarity 46.0%; Pred. No. 2e-06;
Matches 417; Conservative 0; Mismatches 487; Indels 3; Gaps 1;
QY 3 GAGCGGCTTCCCGCGAGCGTGTGTCGTCGCGCGCGAGCGCGCGCGCGCGCGATGTTCCG 62
DB 14731 GCGCGTACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14672
QY 63 CGGCGTGTGTCGCGGAGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 122
DB 14671 GCGGCGCGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14612
QY 123 ACGCGCGGA--CGCGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 179
DB 14611 CGCGGCGGAGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14552
QY 180 CGCGCTCCCGGAGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 239
DB 14551 CGCGCGGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14492
QY 240 CGTGGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 299
DB 14491 GCGCGCGCGCGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14432
QY 300 CGGAGCGGCGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 359
DB 14431 CGAGGAGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14372
QY 360 CAACCGGATGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 419
DB 14371 CTTGCCCGGACCTCTGTGTGCGCGGACCGGCGCGCGCGCGCGCGCGCGCGCGCG 14312
QY 420 CAGGAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 479
DB 14311 GCGGTCGTGTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14252
QY 480 CGTACGCGCTTACCGCGGAGAGCGACGACCGGAGCGCGCGCGCGCGCGCGCGCGCGCG 539
DB 14251 CGAGCGGAGCGCGGTGTGACCTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 14192
QY 540 CGCGCGGTCCTCTCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 599
DB 14191 GCGGAGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14132
```

```
QY 600 GAGCGGAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 659
DB 14131 GTTCCGCGGACCTGGGGTGTGATCGATGACGCGGCGTGTGCGCGCGCGCGCGCGCG 14072
QY 660 CGGCAAGCGCGGAGGTGTACGGGACATCCAGCGGTCCAAACCCCGGCGCGCGCGCG 719
DB 14071 CGGCAAGCGCGGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 14012
QY 720 CGGCGCGTGTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 779
DB 14011 CGCGCGCGTGTGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13952
QY 780 TGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 839
DB 13951 CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13892
QY 840 GAACCTCGAGCGGCGTGTGCGGAGAACTCCGCGCGCTCATGAGACCGAGCTGCGCGCG 899
DB 13891 CAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13832
QY 900 CCAGGAC 906
DB 13831 CCAGGAC 13825
```

RESULT 13

AA61172
ID AA61172 standard; DNA; 4770 BP.

AC AA61172;
XX
XX
XX 22-SEP-2003 (first entry)
XX

DE Actinosynnema pretiosum polyketide synthase (PKS) gene #3.

KW Maytansinoid; ansamitocin; antitumor; polyketide synthase; enzyme; PKS;
KW gene; da.

OS Actinosynnema pretiosum.

PN WO2003045312-A2.

XX 05-JUN-2003.

PF 21-NOV-2002; 2002MO-US037547.

PR 21-NOV-2001; 2001US-0332158P.

XX (UNIW) UNIV WASHINGTON.

PA Flose HG, Yu T, Leistner E;

PI WPI; 2003-493374/46.

XX Novel maytansinoid produced by bacterial host cell transformed with

PT expression vector comprising open reading frame from ansamitocin gene

PT cluster I of Actinosynnema pretiosum, useful as antitumor agent.

PS Disclosure; Page 59-61; 160pp; English.

XX The invention relates to maytansinoid produced by bacterial host cell

CC transformed with expression vector comprising open reading frame from

CC ansamitocin gene cluster I of Actinosynnema pretiosum. Maytansinoid is

CC useful as a potent antitumor agent. The present sequence is A. pretiosum

CC ansamitocin gene cluster I polyketide synthase (PKS) gene

XX Sequence 4770 BP; 485 A; 1947 C; 1809 G; 529 T; 0 U; 0 Other;

SQ

Query Match 11.5%; Score 111.6; DB 8; Length 4770;
Best Local Similarity 47.6%; Pred. No. 4.2e-06;
Matches 395; Conservative 0; Mismatches 429; Indels 6; Gaps 2;

```
QY 126 GCCGAGCCTGCTGCTGAGCGACGTCACCGCCGCGGCGCGACTCGCGCGCCCT 185
    |||||
DB 2790 GCTGCGGCGGCTCTGCGCGCGCGTGAACGCTGCGCAAGAGCGAGCGCGGCTT 2849
    |||||
QY 186 CCGGAGCGGAGCCTGCTCTGCTGCGCTGACAGAGAGCTGCGCTCAAGGCGCTGCG 245
    |||||
DB 2850 GCTGAGCTCTGCGCTGACCGGATCACCTGAGCGC---ACCTGCGCGAACCGCGCCGAGTC 2906
    |||||
QY 246 GCCCGTGAACCCGCTCATGCGGCGCGGCGCGCTGCTGCGCAACCTGCTCGCTCGGAC 305
    |||||
DB 2907 CGCGCGCGCGGCGACCTGAGCTGCTGCTGCTGCGACCGCACCTGCTGAGCGCTGCT 2966
    |||||
QY 306 GGGCATGAGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCTGCGACGCTGAGCGCTCAACCC 365
    |||||
DB 2967 GCCCGGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCGCT 3026
    |||||
QY 366 GATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCGCG 425
    |||||
DB 3027 CCGCGACACCGCGCGCTGCGCGCGCGCTGCTGCGCACCGCGCGCGCGCGCGCGCTGCT 3086
    |||||
QY 426 CCGGCGCGGCGCTGCGCGCGCTGCTGCGCGCGCTGCTGCGAGGCGCGCGCGCGCGCGCT 485
    |||||
DB 3087 CAGGCGCGCTGCG---GACGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCTGCGCG 3143
    |||||
QY 486 GCTCAGCGCGGAGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCGCGCG 545
    |||||
DB 3144 GCTCGGATCTCCCGCGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3203
    |||||
QY 546 GCTCTCTCTCTTGGGCTGCGCGCGCTGCGCGCGCGCTGCGCGCGCTGCGCGCGCGCG 605
    |||||
DB 3204 GGGCGCGCTGCTGCGCGCTGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCG 3263
    |||||
QY 606 GAGGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 665
    |||||
DB 3264 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3323
    |||||
QY 666 CCGCGAGTGTACGCGGAGCATCCAGCGCGCTCAACCGCGCGCGCGCGCGCGCGCGCG 725
    |||||
DB 3324 GAGCGAGGTGAGGACCAAGTTCGCACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3383
    |||||
QY 726 GCTCGCGGAGCGCGCTGCGCTCTTTCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCG 785
    |||||
DB 3384 GCGCGCGGAAACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3443
    |||||
QY 786 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 845
    |||||
DB 3444 CCGCGGAAACCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3503
    |||||
QY 846 CGACGCGCTCTTTCGCGGAACTCCGCGCGCGCTCATGCGGACCGGAGCTGCGCGCGCG 905
    |||||
DB 3504 CCACTGCTGCTCTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3563
    |||||
QY 906 CCACTGCGGAGCTGTTTCGCAACCTTCACCGCGCGCGCGCGCGCGCGCGCGCGCG 955
    |||||
DB 3564 CGACTGAGAGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3613
    |||||

RESULT 14
AAL61173
ID AAL61173 standard; DNA; 9975 BP.
XX
AC AAL61173;
XX
DT 22-SEP-2003 (first entry)
XX
DB Actinobynema pretiosum polyketide synthase (PKS) gene #4.
XX
KM Maytansinoid; ansamitocin; antitumour; polyketide synthase; enzyme; PKS;
XX gene; ds.
XX Actinobynema pretiosum.
XX
PN WO2003045312-A2.
```

```
XX 05-JUN-2003.
PD
XX 21-NOV-2002; 2002WO-US037547.
PF
XX 21-NOV-2001; 2001US-0332158P.
PR
XX (UNIW ) UNIV WASHINGTON.
PA
XX Floss HG, Yu T, Leistner E;
XX
XX WPI; 2003-49374/46.
DR
XX
XX Novel maytansinoid produced by bacterial host cell transformed with
PT expression vector comprising open reading frame from ansamitocin gene
PT cluster I of Actinobynema pretiosum, useful as antitumor agent.
XX
XX Disclosure; Page 62-67; 160pp; English.
XX
XX The invention relates to maytansinoid produced by bacterial host cell
CC transformed with expression vector comprising open reading frame from
CC ansamitocin gene cluster I of Actinobynema pretiosum. Maytansinoid is
CC useful as a potent antitumour agent. The present sequence is A. pretiosum
CC ansamitocin gene cluster I polyketide synthase (PKS) gene
SQ
Sequence 9975 BP; 947 A; 4102 C; 3795 G; 1131 T; 0 U; 0 Other;
```

```
Query Match 11.3%; Score 109.8; DB 8; Length 9975;
Best Local Similarity 46.8%; Pred. No. 5,9e-06;
Matches 414; Conservative 0; Mismatches 467; Indels 4; Gaps 2;
```

```
QY 69 GCTGCGGAGGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 128
    |||||
DB 3716 GCGGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3775
    |||||
QY 129 GAGCGCGCTGCC-TGATGAGCGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 187
    |||||
DB 3776 CCGCGCGAGCGCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3835
    |||||
QY 188 GAGAGCGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 247
    |||||
DB 3836 ACGACTGATCACCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3895
    |||||
QY 248 CCGTGAACCGGAGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 307
    |||||
DB 3896 GGGTCAACCGCGCGAGATTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3955
    |||||
QY 308 GCATGCGCGGAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 367
    |||||
DB 3956 CCGAGGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 4015
    |||||
QY 368 TGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 427
    |||||
DB 4016 ACCTGTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4075
    |||||
QY 428 GCGCGGAGCTCAAGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 487
    |||||
DB 4076 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4135
    |||||
QY 488 TCACGCGGAGGAGCAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 544
    |||||
DB 4136 GCGCGCTGCACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4195
    |||||
QY 545 TGTCTCTCTCTTTCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 604
    |||||
DB 4196 CGCACCGCTCAACCGCGCGCTGTCACCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 4255
    |||||
QY 605 CGACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 664
    |||||
DB 4256 CCGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4315
    |||||
QY 665 GCCCGAGGTGTACGCGGAGCATCCAGCGGTCAACCGCGCGCGCGCGCGCGCGCGCGCG 724
    |||||
```

```

Db      4316 TGCAGAGCTGACCTCCGACCTGACGCTGTTTCGCTCTTCCTCGACCTCCGAGCTGC 4375
QY      725 CGCTCCGCGAGGCGCTGCTGCTCTTCGCGCCGCGCTGAGCGAGCGACCGACCGTGGCG 784
Db      4376 TCGGCGCGCGGCGGCGGCGCACTACGCGGCGCGCAAGCTTCTCGAGCGGCTGAGCC 4435
QY      785 ACGCGCCCGGCGCGCGCGACGCGCCCGGCGCATCCGCGGAGTGCAGCGGCGCGGAAAC 844
Db      4436 GCGAGCGCGACGCGCGCGGCGCTCCGCGCACCTGATCGCTGGGCGCTGTGGGCGGAGG 4495
QY      845 TCGACGCGCTCTTCGCGGAACTCCGCGGCTCATGGAGCGGAGCTGCGCGCGGCGCAG 904
Db      4496 CGACGCGCTCACCGGCTCCCTGTCCGAGCGACCGGAGCGCATGCGCGCTCCGCGG 4555
QY      905 ACCACTGCCAGAGCTGTTCCGACCTCCACCGCACCGAGAG 949
Db      4556 TCGCGCCCTTGACCGGAGCGGCGCTCCGCTGTTTCAGCGCG 4600

RESULT 15
ID      AAD54645 standard; DNA; 113193 BP.
XX
XX      AAD54645;
XX
XX      26-JUN-2003 (first entry)
XX
DE      Streptomycetes nodosus amphotericin (amph) biosynthetic gene cluster.
XX
KM      Polyene; antibiotic; amphotericin; amph; polyketide; enzyme; gene; ds.
XX
OS      Streptomycetes nodosus.
XX
FH      Location/Qualifiers
CDS     complement(4..1824)
        /tag= a
        /product= "ABC transporter encoded by S. nodosus amphG
        gene"
        complement(1805..3628)
        /tag= b
        /product= "ABC transporter encoded by S. nodosus amphH
        gene"
        /tag= .4874
        /tag= c
        /product= "GDP-mannose dehydratase encoded by S. nodosus
        amhdIII gene"
        5042..33574
        /tag= d
        /product= "Polyketide synthase multienzyme housing
        extension modules 9, 10, 11, 12, 13 and 14 encoded by S.
        nodosus amph gene"
        33584..50518
        /tag= e
        /product= "Polyketide synthase multienzyme housing
        extension modules 15, 16 and 17 encoded by S. nodosus
        amphJ gene"
        50571..56675
        /tag= f
        /product= "Polyketide synthase multienzyme housing
        extension modules 18 and thioesterase encoded by S.
        nodosus amphK gene"
        56829..58019
        /tag= g
        /product= "Cytochrome P450 encoded by S. nodosus amphL
        gene"
        58139..58648
        /tag= h
        /product= "ORF1, hypothetical protein"
        complement(58756..59610)
        /tag= i
        /product= "ORF2, hypothetical protein"
        59869..61470
        /tag= j
        /tag= j

```

```

FT      FT
FT      /product= "ORF3, hypothetical protein"
FT      /transl_except= (pos:59869..59871, aa:Met)
FT      CDS complement(61798..61995)
FT      /tag= k
FT      /product= "Ferredoxin encoded by S. nodosus amphM gene"
FT      complement(62051..63250)
FT      /tag= l
FT      /product= "Cytochrome P450 encoded by S. nodosus amphN
FT      gene"
FT      complement(63250..64308)
FT      /tag= m
FT      /product= "NDP-sugar aminotransferase encoded by S.
FT      nodosus amhdIII gene"
FT      CDS complement(64324..65775)
FT      /tag= n
FT      /product= "Glycosyl transferase encoded by S. nodosus
FT      amhdI gene"
FT      /transl_except= (pos:65773..65775, aa:Met)
FT      CDS 66081..70319
FT      /tag= o
FT      /product= "Polyketide synthase multienzyme housing
FT      loading module encoded by S. nodosus amphA gene"
FT      70366..79938
FT      /tag= p
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 1 and 2 encoded by S. nodosus by amphB
FT      gene"
FT      79956..112709
FT      /tag= q
FT      /product= "Polyketide synthase multienzyme housing
FT      extension modules 3, 4, 5, 6, 7 and 8 encoded by S.
FT      nodosus by amphC gene"
XX
XX      WO200297082-A2.
XX
XX      05-DEC-2002.
XX
XX      27-MAY-2002; 2002WO-1E000071.
XX
XX      31-MAY-2001; 2001IE-00000527.
XX
XX      (UYDU-) UNIV COLLEGE DUBLIN.
XX
XX      Caffrey JP;
XX
XX      WPI: 2003-201271/19.
XX      P-PSDB; AAE36116, AAE36117, AAE36118, AAE36119, AAE36120, AAE36121,
XX      AAE36122, AAE36123, AAE36124, AAE36125, AAE36126, AAE36127, AAE36128,
XX      AAE36129, AAE36130, AAE36131, AAE36132.
XX
XX      Novel cytochrome P450 enzyme and nucleotides encoding the enzyme, useful
XX      for preparing amphotericin derivative or analog antibiotic agent with
XX      altered properties, in biosynthesis of polyketide other than
XX      amphotericin.
XX
XX      Claim 1; Page 52-114; 276pp; English.
XX
XX      The invention relates to the gene cluster encoding the polypeptides
XX      responsible for the biosynthesis of the polyene antibiotic amphotericin
XX      (amph) of Streptomycetes nodosus. Polynucleotides of the invention are
XX      useful for preparing amphotericin derivatives or analogue antibiotic
XX      agents with altered properties and in the biosynthesis of polyketides
XX      other than amphotericin. amhdIII, amhdII or amhdI mutants are useful
XX      for producing amphotericin derivatives glycosylated with alternative
XX      sugars; amhdIII or amhdII gene sequences are useful in engineered
XX      biosynthesis of perosaminyl-amphoteronolide B; amhdIII or amhdII and
XX      amhdI gene sequences are useful in the engineered biosynthesis of
XX      perosaminyl-16-desacetoxy-16-methyl amphoteronolide B; amhdIII, amhdII
XX      and amhdI gene sequences are useful for preparing polypeptides capable
XX      of addition of mycosamine to a polyketide other than amphoteronolide A or
XX      B or for preparing polypeptides for in vitro synthesis of GDP-mycosamine.
XX      The present sequence is S. nodosus amph biosynthetic gene cluster
XX

```


SQ Sequence 113193 BP; 14248 A; 45141 C; 38354 G; 15450 T; 0 U; 0 Other;

Query Match	Score	DB	Length
11.2%	108.8	8	113193

Best Local Similarity 46.2%; Pred. No. 4.3e-06;

Matches 439; Conservative 0; Mismatches 502; Indels 9; Gaps 2;

QY	6	CGGCTTCCCCCGCAGCGCTCGTCCGCGGCGACCGGGGGGGTGGGGGCAATGTTCCCGG	65
Db	83294	CGCGGCTTACAGACCTTCGGGCGCATCAGCCGGGGGGGCTGCCGTTGCGCTGGAGGG	83353
QY	66	GCTGCTGCGGAGACCGGGCAGCGCAGCGCTGTCGTGACCTTGTAACCGCGCGCGAGC	125
Db	83354	CGTGCACCTGCGCGCGTCCGGCGCACACCGTCCCGCGCGGATACCTCGTGGCGCA	83413
QY	126	GCCGAGACCTGCTGCTGTGGAGCACTCACCGCGCGCGGAGCCGAACTGCGCGCGCCT	185
Db	83414	GGAACCGTTCATCCGCTGTATACACACCGGGGGCGGACCGTCTCTCGTGGATTC	83473
QY	186	CGGGAGCGGGAAGCTGCTGCTGTCGCGGTACAGAGAGCGTGGCCTCAAGGCGGTGC	245
Db	83474	CTGTGTTTCCCGGCGGCTCCCAAGAGATCCCCGGACCGCGGGGCGTCCACCGCA	83533
QY	246	GCCCGTGAACCCGCGCTATGCGCGCGGAGCGGCTGTGCGCGACACCTGTCCGTCCGAC	305
Db	83534	CTCCCTTTCAGGTGACTGAC-----CCGCTGGGAGGCACTCGGGGCACTCCCGA	83590
QY	306	GGGCAATGGCCGCGGAGCTCGGGGCCACGCCCCGCGCTTCAGACATGAGGCTCAACC	365
Db	83591	GACCTGCGCCTCGTGGCGCGCCGACCGGCGCGGACTGCGCGACACCTCGCACCGGCGG	83650
QY	366	GATGTTGCCCCCGCGCGCGGATGACCGGCGGACCCGTGGCGCGCGTGTCAACAGGA	425
Db	83651	CTTCAGATACCGTCCACCCGACCTGGCGGACTGTGGCGCGCGGACCCGCTCGGTGC	83710
QY	426	CGGGCGGCGGTACAGGCTCTGTGCGGCTGTGCGAGGCGGCGGCGGACCGCTAGC	485
Db	83711	CGGGACCGTGTACACGCGCTTGGCGGGCACAGACACCCGCGCGCGGACCGCCCA	83770
QY	486	GTTCACGGCGGAGAGACGACCCGGAAGACGCGGCGCACCCAGGCTTGACGACGCGT	545
Db	83771	CACACCGGACGCGCGTACTCGGCTGGGACAG-----CACTGCTGACGAGGACCG	83824
QY	546	GCTCTCTCTTGGGGCTGCGCCCTCGCGCGCGCTCGCGGCTGTGACGTTCCGGGCTTGCGGC	605
Db	83825	GTTCGCGGAGCGCCGCGTGTCTTGTGTACCTCGACCGCACCGCGGCGGTGACCTTGC	83884
QY	606	GACGAGACCGCGCGCCACACAGGTGTGACTCGGCTTGTGCGCCGCTGTGCGCGGACG	665
Db	83885	GGCGCGCGCGGCGCGGCGTGTATCCGCAACCGCCGACCGAGAACCCCGGCGCTTGT	83944
QY	666	CCCCGAGGTATCGGGGACATTCAGCGGTTCACCCCGGCGGCGGCTTCGCGCGCGCGC	725
Db	83945	CTCTCTGACCTGCGCACCGGACCCCGCGCGGCGCGCGGCGCGCTGTGGCGGCTTGC	84004
QY	726	GCTCGCGAGAGCCCTCGCTCTTTCGCGCGGCTGTGTGGCGAGACCGCGACCTGTGCCA	785
Db	84005	CGCGCGGTATGAGAGAACCCGACTTCGCGGTACCGGCTTCGAGTACTATACCGCGCGCT	84064
QY	786	CGCCCCCGGCGCGCGACGCGCCCGGAGCAATCCCGGAGGATGAGACGGCGCGGGAAGCT	845
Db	84065	CGCGCGGAGTCCGCGCACCGACCGCGCGGCGCGGCTGAAACCGGAGCGGACCTGTCTGAT	84124
QY	846	GACGCGCGTTCGAGGAACTTCGCGCGGCTCATGAGACCGGAGCTTCGCGGCGGCGCAGA	905
Db	84125	CACCGGAGGACGCGGAGACTCGGCGGCGGCTTCGCGCGGACCTGTCAACGGAACGGG	84184
QY	906	CAACTGCGAGGAGCTGTTCCGACCTTCACCGCACCGACGAGGAGGCG	955
Db	84185	CGTTCGCGCGGCTGTCTGTGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG	84234

This Page Blank (uspio)


```
Db 259 GTCTGAGCGGGAGTATGACGGCCCGGTCGCTGCGGACACTTGTGCTCAAGAGC 318
Qy 307 GGCATGCGCCGAGAGCTCGCGGCCACGCCCGCGAGCTCAGCAGTGGAGCTCAACCCG 366
Db 319 CGGATGCGCGGGGGGCTGAGAGGGGCGCCCGGGGTGACAGGGGCTGAGACCCG 378
Qy 367 ATGTTCGCCCCCGCGCGGAGTACCGGCTGACCCGTGGCCCGCTGTGTCACCAAGGAC 426
Db 379 ATGTTCGCCCCCTGCTGAGGTCTCAAGGGCGCGCGGTGGCGCGGTGTGTACCGAC 438
Qy 427 GGGCGGGGCGCTCACGGCGCTGCTGCGGCTGTGAGAGGGCGCGCGCGCGCGAGCCGTAACG 486
Db 439 GGGCGCGGTGTGCGGCGCTGTGTGAGCTGTGTGCGCGAGTGGAGAGCCGAGTGTGAG 498
Qy 487 CTCACGCGCGAGAGACGACGACCGGACGACCGCGCGCACCGAGCCCTGACGACGCGCTG 546
Db 499 ATCCGCGCGCGGGGACGACGAGCTGACCGCGCGGACGAGAGCCCGCACGCAATGCCGCG 558
Qy 547 CTCTCTCTCTTGGGAGCTGCGCTTGGCGCGCGCTGCGAGCTGACGTCGCGAGCCCTGCGCGCG 606
Db 559 GTCTGAGCTTTCGGGGCTGGGCGCTGGGTGAGCTGTGCGGTGACGTTGGGGGCGCTGCGGGAC 618
Qy 607 AGCGGACCGCGCGCCACGACGAGTGTGCTGCGCCCTCTGAGCCGTGTGCTGCGCGCGACG 666
Db 619 AGTCCCGCGCGCGCGCTGTGAGCTGTGAGCTGTGCGCTGTGCGAGTGTGCGCGAGACG 678
Qy 667 CCCGAGGTGACGAGGACATCAGCGGTCCAACTCCCGGCGCGCGCTGCGCGCGCGCGCG 726
Db 679 CCGGAGGTGATTTTCAGACATCCAGGCGCGCACCTCCGCGCGCGCGCGCGCGCGCGCG 738
Qy 727 CTGCGCGAGGCGCTGCGCTCTTTCGCGCGCGCTGTGCGAGCACCGCGACCGTGTCCGAC 786
Db 739 CTGCGCGCGCGCGCTGTGCTGCGCGCGCTGTGCGAGCGAGCGAGCACGAGACGTTTC 798
Qy 787 GCCCGCGGCGCGCGCGACCGCGCGCGCGCAATCCGCGGAGACGACG 832
Db 799 GCCGCTGTTCGCGGACATGCGCGGTGTGTGCGAGCAGCAGCGCG 844
```

RESULT 2

```
US-09-987-614A-2
; Sequence 2, Application US/09987614A
; Patent No. 683382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMBAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
; US-09-987-614A-2
```

Query Match 27.9%; Score 270.8; DB 4; Length 888;
Best Local Similarity 58.8%; Pred. No. 9.3e-34;
Matches 486; Conservative 0; Mismatches 337; Indels 3; Gaps 1;

```
Qy 10 TTCCCGGAGCGTCTGTCTGCGCGGACGCGGCGGTGGAGCGCATGTT---CGCCGGG 66
Db 19 TTCCGAGCGTTGTGTGTGTGTGGCGGGCGGGTGTGGGCGCATGTTTACGCACTGG 78
Qy 67 CTCTGCGGAGAGCGGCGAGCCGACAGCTGTCTGTGTGACCTGTGACCGCGCGGAGACG 126
Db 79 CTGTGTGCTTTCGAGGGGTGGCGGTGACCTGTGAGCTGTGACGCGGCGCGGTGTGAGAC 138
Qy 127 CCGGAGCGCTGCTGTGTGAGCGACGTCACGCGCGGAGCGCGAACTGCGGCGCGCTTC 186
Db 139 GGGGTGCGGGGTGTGTGTGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 198
Qy 187 CCGGAGCGGAGACTGTCTGTCTGTGCGCGTACAGAGACGTGAGCTTCAAGGCGGTGAG 246
Db 199 GCGGCGGCGGAGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 258
Qy 247 CCGGTGACCGCGCTCATGTGCGCGCGCGCGCGCTGTGTGTGTGTGTGTGTGTGTGTGT 306
Db 259 GTCTGTGCGGGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 318
Qy 307 GGCATGCGCGGAGAGCTCGCGGCCACGCGCCCGGAGTTCAGACGTTGAGCTCAACCCG 366
Db 319 CGGATGCGCGGCGGAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 378
Qy 367 ATGTTCGCCCCCGCGCGCATGACCGGCGACCGCGTGTGCGCGCGCTGACGAGGAC 426
Db 379 ATGTTCGCCCCCTGCTGTGTGTGTCTTTCAGGGGCGCGCGGTGTGTGTGTGTGTGTGT 438
Qy 427 GGGCGGGGAGTACAGGCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 486
Db 439 GGGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 498
Qy 487 CTCACGCGGAGAGACGACGAGCGAGCGCGCACCGAGCGCGCTGTGACGACGCGCGT 546
Db 499 ATCCGCGCGCGCGCGGACGACGAGCTGTGACCGCGCGCGAGCGCGCGCGCGCGCG 558
Qy 547 CTCTCTCTCTTGGGAGCTGCGCTTGGCGCGCGCTGTGAGCGTGTGAGCGTGTGCGCG 606
Db 559 GTCTGAGCTTTCGGGGCTGGGCGCTGGGTGAGCTGTGAGTGTGTGTGTGTGTGTGTGT 618
Qy 607 AGCGGACCGCGCGCCACGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 666
Db 619 AGTCCCGCGCGCGCGCTGTGAGCTGTGAGCTGTGCGCTGTGCGAGTGTGCGCGAGAC 678
Qy 667 CCCGAGGTGACGAGGACATCAGCGGTCCAACTCCCGGCGCGCGCTGCGCGCGCGCGCG 726
Db 679 CCGGAGGTGATTTTCAGACATCCAGGCGCGCACCTCCGCGCGCGCGCGCGCGCGCG 738
Qy 727 CTGCGCGAGGCGCTGCGCTCTTTCGCGCGCGCTGTGCGGACGACCGCGACCGTGTCCGAC 786
Db 739 CTGCGCGCGCGCGCTGTGTGCGCTGTGGGGCAGCGCGTGTGAGAGGGGCGAGAGACGTTTC 798
Qy 787 GCCCGCGGCGCGCGGACCGCGCGCGCGCAATCCGCGGAGACGACG 832
Db 799 GCCGCTGTTCGCGGACATGCGCGGTGTGTGCGAGCAGCAGCGCG 844
```

RESULT 3

```
US-08-765-907A-1/c
; Sequence 1, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMBAS-JACOUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
```

```

? TITLE OF INVENTION: Mutasesynthesis
? FILE REFERENCE: Streptogramin genes
? CURRENT APPLICATION NUMBER: US/08/765, 907A
? CURRENT FILING DATE: 1997-03-20
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn Ver. 2.0
? SEQ ID NO: 1
? LENGTH: 2888
? TYPE: DNA
? ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

```

Query Match	27.9%	Score 270.8	DB 3	Length 2888
Best Local Similarity	58.8%	Pred. No. 8.3e-34		
Matches 486; Conservative	0	Mismatches 337	Indels 3	Gaps 1

oy 10 TTTCCCCAGCGTGTGTGGGCGGACGGGGCGTGTGGCGCATGT---CGCCGG 66
 Db 1818 TTTCCGCGTGTGTGTGGGCGGCGCCGCTGTCGGTGGCCGCGATGTTCAGCCACTGG 1755

QY 67 CTCTGCGGGAGGCGGGCAGCCGACTGTCGTCTGCACTCTGTACCGCGCCCGGGACGG 126
Dδ 1758 CTGATGCGTTGCGGGGCTGCGCGGTACCTGAGCTGAGCCCGGGCGCGGTGCGCGCGGAC 1699

127 CCGGACGCTGCTGGTGGTGGGAGACATCACCGGCGGGGCGGACTCGCGGCGCGCCCTC 186
 1698 GGGGTGCGGGTGTGGCCGGGTGATGTGCGGCGGCGGGGCGGAGAGCGGCTGCGCGGCGCTG 1639

Db 1638 GCGGCGGCGACGTGTGTGCTGCTGCGGAGCCGGTGGCGTGGAGGCGGTGAG 1579

Dy 287 CCCCAGACCGCATGAGCCCGGGCCTGCATGCCACATCTGTCCGTCGGAGC 306

Dd 1578 GTCTCGCGGGGTGATCGGGCCCGTGCCTGCTCGCGGACACTTGTCGGTCAAGAC 1519

D_b

1518 CGATGCGCGGGCGAGCTGTGAGCGCGCGCCGGGGCTGCAGGCGGTGGGGCTGAACCG 1459

Db 1458 ATGTTGCCCCCTGCTGAGTCTTTCAGGGAGCGCGCGGTGCTGATCCGAC 1399

Db 1398 GGGCCCGGTGTGCGGGCCCTGTGTGAGCTGGTGGCCGGGTGGGGGGCCCGGGGTGTGAG 1399

1338 ATGCCGCGCGCGGCAACGAGCTGACCGCCGCGCAGCAGGCCCGCCACGCAATGCCGCG 1279

Db 1278 GTCCTGGCCTTCCGAGCTGGGCGCTGAGCTGCGGTGACGTCGGGCGCTCCGGGAC 1219

Db 1218 AGTGGCCCCCGCCGCATCTGGCGCATCTGGCGCTGCTGCCCCGATCGCCGGCGGACG 1159

Db 1158 CCGAGAGTGTATTTGACATCCAGGCCGCCAATCCCCGGGCGCGCGGCGCGGCGGCGGCGG 1099

Db 1098 CTGGGCGCGGCTGTGCGGCTGGGCGAGCCGTCGAGAGGGCGACGAGAGAGCTTC 10399

Db 1038 GCCGCCCCCTGTTCCGCCGAACTGCGCGGTGTCTGGGCGAGCAGCGCG 953

```

RESULT 4
; Sequence 1, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: TRIBOUT, Denis
; APPLICANT: BAMBAS-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBOUSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSER, Gilles
; TITLE OF INVENTION: Metaprogams And Method For Preparing Same By
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/09/987,614A
; CURRENT FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US/08/765,907
; PRIOR FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-1

```

Query Match	27.9%	Score 270.8	DB 4	Length 2888
Best Local Similarity	58.8%	Pred. No. 8.3e-34		
Matches 486	Conservative 0	Mismatches 337	Indels 3	Gaps 1

oy	10	TTTCCC	CGCAG	CGTCGT	CGGCGG	CGACG	GGGGCG	CGTGGG	CGGCGAT	GT	----	CGCCGG	66
Db	1818	TTCCGG	CGCTGT	GTGTGT	GCTGGG	CGGGCG	CGGCGG	CGTGGG	CGGCGAT	GT	CGCC	CACTGG	1755

67 CTGCTGCGGAGGCGGCGAGCGCCACGCTGTGTGCACCTCTGATCCGCGCGCGCGGACGG 126
 Oy |||||
 1758 CTGCTGCTTCGCGGGGCTGCGCGGATCTGGCTGACCTGCGCGGGCGCGGTGCGCGGAC 1699
 Db |||||

QY 127 CCGGACGCTGCCTGATGGCGACGTACCGCGCGCGGGGACCCGAACTCGCGCGCGCCCTC 186
 Db 1698 GGGGTGCGGGTGTGTCGCCGTGATGTCTCGCGCGCGCCGGGGCGGAGGCGGTTCGCGCGCGCTG 1639

Db 1638 GCGGCGGCGACGTGTGTGTGTGCGGAGCCGGTGGCGTGGAGGCGGTGAG 1571

247 CCCGAGACCCGATCATCCGCGCGCGCGCGATCTCTGCGGACATCTGCTCCGCGGAC 308
Db 1578 GTGCTGGCGGGGTGATGCGCGCGCGCGGTGCTCTCGCGGACACTTGTGCGTCAAGAC 1519

27 507 CGATCGCGGCGCTGCGTGGCGCGCGCGGCGCTGCAGCGCGTGGCGCTGAACCG 1459

Db

1458 ATGTTGCCCCCTCGTGGGTTTCAGGGGGCGGCGGCGGGTGGTGCACCGAC 1399

Db 1398 GGGCCCGGTGTGCGGCGCCCTGTGAGACTTGTGTGCGGAGTGGGGGGCGCCCGGGTGTGAG 1339

Db
1338 ATGCCGCGCGCGCGCGCACGACGAGCTGACCGCGCGCGCAGCAGGCGCCCGCCACGCAATGCCGCG 1279

Db 1278 GTGCTGGCCTTCGGGCTGGGCTGAGACTTCGGTGGACGTGGGCGGCTGCCGGAC 1219

QY	607	CCGCGACCGCGCGCCACCAAGATGTGTGCGCCCTCGAGGCGAGTGCCTGAGGCGAGCG	666
Db	1218	AGTGCCTCCGCGCGCCGCATCTGAGCATGTGAGGAGCTGCTGTGAGCCCGAATGCCGCGAGACG	1159
QY	667	CCCGAGGTGTATCGGAGGACATCCAGCGAGTGCACAACCCCGAGGCGAGGTCGCGCGCGGAGCG	726
Db	1158	CCGAGGAGTATTTTCACATCCAGGCGCGCAACCCCGCGCGCGCGCGCGCGCGGCGAGGCG	1039
QY	727	CTCGCCGAGGCGCCCTGCGCTCCCTTGCGCGCGCGCTGATCGCGACGACCCCGGACCGTGCAGC	786
Db	1098	CTGAGCGCGCGCGCTGTGTGCGCGCTGAGGCGAGGCGCGCTGTGAGAGAGGCGACGAGAGACGTTTC	1039
QY	787	GCCTCCGAGGCGCGCGGACGCGCCCGCGGCCATCCGCGGAGATGCGGAC	832
Db	1038	GCGCGCTCTTTTCGCCGAACTGCGCGGAGTGTCTGAGCGAGCAAGGCG	993

```

US-RESULT 5
US-0804-804-227C-7/c
Sequence 7, Application US/0804227C
Patent No. 5876991
GENERAL INFORMATION:
APPLICANT: Dehoff, Bradley S.
APPLICANT: Kuhnloss, Stuart A.
APPLICANT: Rostreck, Paul R., Jr.
APPLICANT: Suton, Kimberly L.
TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: THOMAS G. PLANT 1501
STREET: LILLY CORPORATE CENTER
CITY: INDIANAPOLIS
STATE: IN
COUNTRY: USA
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: ASCII (DOS) Text only
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/804,227C
FILING DATE: February 21, 1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Plant, Thomas, G.
REGISTRATION NUMBER: 35,784
REFERENCE/DOCKET NUMBER: X-8231
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-276-2459
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 44377 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 350..14002
FEATURE:
NAME/KEY: CDS
LOCATION: 14046..20036
FEATURE:
NAME/KEY: CDS
LOCATION: 20110..31284
FEATURE:
NAME/KEY: CDS
LOCATION: 31329..36071
FEATURE:
NAME/KEY: CDS
LOCATION: 36155..41830

```

US-08-804-227C-7

Query Match	10.8%;	Score 104.4;	DB 2;	Length 44377;
Best Local Similarity	46.7%;	Pred. No. 2.9e-08;		
Matches 438; Conservative	0;	Mismatches 491;	Indels 9;	Gaps 3;

QY	22	GTGTCGTCTCGGCGGACAGCGGGGCGGATGGAGCATTTTGTGCGCGGCTGTGCGGAGGGG	81
Db	29490	GTGTCGTGTGATATCCATATGCGCGGACGGGCGTCCACTTGTCCGGGGCTTCGGGTGGCCAG	29431
QY	82	GGAGCGCGACGCTGTCTGTGCACCTGTATCCGCGCGCGGAGACGGCGGACGCTTGCCCTG	141
Db	29430	CACCTTCGCGC---CGAGGTGGCGGGCGATTTTGCACCGCGGCATATGCGCACGCCCGGAC	29374
QY	142	GTGGGAGACGTATCACCGCGCCGGGGGCTCCGAATTCGCGGCGCCCTCCGGAGCGCGAATCTC	201
Db	29373	GGCGGCGTGCACAGACACCCGCTCTCCCGAGCCTGAGGCGGGCCAGGGTATACAGACCTGTAT	29314
QY	202	GTCTGTGTCGCGCTTAC---CGAGAGCATGTGGCCCTCAAGGCGGTGGCGCGCCGTGACCGG	258
Db	29313	GTATGGCGCTATGAGAACACCGCGGGGAGCCGAGCGCGCTGTGGGTATGATCAACGCTTCGG	29254
QY	259	CTCATGTGCGCGGCGCGCGCTCTCTGCGGACACCTGTCTCCGTCCGAGCGGAGCATGGC---C	315
Db	29253	GACGCGGGGCGAGCAGCGCGGGGTGTGGCGCACGCAAGCGGGGCCAGGGCGCGCTCCACAG	29194
QY	316	GCGGAGCTTGTGCGGCGCCACGCGCCCGCGCGTTCAGACATGTGGGCTTACCTCCATGTTCCGC	375
Db	29193	GCGGAGCACCGGTGTCCCGACGCGCGACACCGGTATGCGCGGGGCGGACCTTCAGACACGAT	29134
QY	376	CGCGCGCGCGGACATGACCGGCGGACCCGCTGTGGCGCGCGGTATGATCAACAGGAGCGGGCGGGG	435
Db	29133	GCGGGGCGCTCTCGTTTCCATCTTCGCGCTTCGCGCGGGGTATCATGCGGAGGCGATGAGGGC	29074
QY	436	GTCACGCGCCCTGTCTGTGCGGCTGTGTGAGGGCGGCGGCGGAGCGCGGACCGGTACGCTCATCGGCG	495
Db	29073	GTCGCGGAGATTGACCGCGCGCGCGGTATCGGGGAGACGCGCACCTGAGCGGGCTTCAGGGG	29014
QY	496	GAGGAGACGACCTCGGACGACGCGGCGCCACCCAGGCGCTTATGCGACGCGGTCTCTTCTCC	555
Db	29013	CGCTTCGGGTGTCCGGGCGGGGACACAGGCGAGGCGCTGTGAGGGTATGATCTGATGCCCGG	28954
QY	556	TTTGGGCTGTGCGCTTCGCGCGCGCTCGCGCGTTCGACGATTCGGGCGCTTGGCGCGACCGGACCG	615
Db	28953	GCCCAAGCGCAGGACACCGTCTGTGGCGGGGTATGAGCGGCGCTCTCGTTCGGTTCGCGCGC	28894
QY	616	CGCGCCCAACAGGTGTGCTGTGCGCCCTCTGTGCGCGGTATGTCTGCGCGGAGCACCCGAGGTG	675
Db	28893	CGCTCTCAGCGCGGGGACGATACACGCTCCCGGCGGAGACCGGCGATCTGGGTTCGCCGGA	28834
QY	676	TATGGGAGCATTCAGCGGTTCACACCCCGGGCGGGCTTCGCGCGCGCGGCGCTTCGCCAG	735
Db	28833	GGGAGAGGGGTTCGATAGCGCGCCCGGCGGACCGCGGATGTATGTCAAGTTCACAGAGAG	28774
QY	736	GGCCTTACGCTTTCGCGCGCGCTGTGCGCGACGACACCGGATCCGTGCGGAGCGCGCCCGGAG	795
Db	28773	GAAACCGCGCGGCTCTTCGCGCTGTGGCGGAGCGCAGAGGCCCCACGCGCGGCGGCGCGGAC	28714
QY	796	CGCGCGAGCGCGCCCGCGGCATCCGAGGGGATTCGACGCGCGCGGAGAACCTTCGACGCGCTC	855
Db	28713	CGGGTGTTCACACTCTCGGCGGTGTGTGCGGTATCGGCGACCGGCGCGGGGTGTGCGCACACAG	28654
QY	856	TTTCGGGAACTTCGCGCGGCTCATGTGAGACCGAGCTTCGCGCGGGGCGAGGACCATGTCCAG	915
Db	28653	GCGGGCGCGGTTCGCGGCGCTTCGCGGGGCGCGAGCGCGGCGGTGAGAGGATTCAGACACCG	28594
QY	916	GAGCTTTCGACACCTTCACCGGACCGACCGACGAAAG	953
Db	28593	GCAAGCCAGGTGTGTGCGGTGTGCGCGGACGATGTGCGG	28556

RESULT 6
US-08-804-198-1/c

```

Sequence 1 Application US/08804198
: Patent No. 5945320
: GENERAL INFORMATION:
: APPLICANT: Burgett, Stanley G.
: APPLICANT: Kuheloss, Stuart A.
: APPLICANT: Rao, Nagaraja R.
: APPLICANT: Richardson, Mark A.
: APPLICANT: Roestek, Paul R., Jr.
: TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: PAUL R. CANTRELL 1138
: STREET: LILLY CORPORATE CENTER
: CITY: INDIANAPOLIS
: STATE: IN
: COUNTRY: USA
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: Macintosh
: OPERATING SYSTEM: Macintosh 7.0
: SOFTWARE: Microsoft Word 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/804,198
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: CANTRELL, PAUL R.
: REGISTRATION NUMBER: 36,470
: REFERENCE/DOCKET NUMBER: P9113
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 317-276-3885
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 44377 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 350..14002
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 14046..20036
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 20110..31284
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 31329..36071
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 36155..41830
: US-08-804-198-1
:
: Query Match 10.8%; Score 104.4; DB 2; Length 44377;
: Best Local Similarity 46.7%; Pred. No. 2,9e-08;
: Matches 438; Conservative 0; Mismatches 491; Indels 9; Gaps 3;
:
: 22 GTCGTCGTGCGCGGACGCGGCGCGGTGGGCGGAGATGTCGCGGCGTGTGCGGAGAGCG 81
: 29490 GTCGTGCGTGTGCGGATGCGGATGCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 29431
:
: 82 GGCAGCCGCGACGCTGCTGCTGCTGACCTGTACCGCGCGCGGAGACGCGCGACGCTGCGT 141
: 29430 CACCTCCGCGC---CGAGGTGGCGGCGGATCTGCACCGCGGCATGCGGACGCGCGCGCG 29374
:
: 142 GTGGGCGACGTCACCGCGCGCGGCGCGGAACTGCGGCGCGCGCTCCGCGGAGCGGAGACTTC 201
: 29373 GCGCGCGTGCACGAGCACCGCTCCCGCGCGCTGAGGCGCGGCGGCGGAGGTGACGAGACTGTA 29314
:
: 202 GTCTGTGCTGCGCGGTACA---CGAGAGGTGAGCCTCAAGGCGGTGGCGCGCGCGGTGACCGCG 258

```

Db	2913	GTAGGCGCTGAGGAACAACCGCGGGGACCGAGCGCGCTGGGGGTAGAGACACACCGCTCCGG	2925
Oy	259	CTCATGAGCGCGGAGCGCGCTGCTCGCGACACCTCTGTCGTCCGACGGGCAATGC--C	315
Db	29253	GACGGGGGCGAGGACCGCGGTGGTCGGCCACGACAGCGGGGCCAGGCGCGCTCCACAG	29194
Oy	316	GCGGAGCTCGCGGCGCCACGCGCCCGCGGTCACAGCAGTGGGCTCAACCCGATGTTCCG	375
Db	29193	GCGAGGACCGCGGTCCCGACCGCGGACACCGGTGACCGCGGGGCGGACTTCGACGACGT	29134
Oy	376	CCGCGCGCGGAGTGAACCGGCGGACCCGCTGCGCGCGTGTGACACAGGGAGGGGCGGGG	435
Db	29133	GCGGCGCGCTCTCGTTCCATCTCGCGCTCGCGCGGGGTACATGCGAGGGGCGATGAGGG	29074
Oy	436	GTCACGGCGCTGCTGCGGCTCGTCAAGAGCGGCGGCGGACGGCGCTACGCTCACGG	495
Db	29073	GTCGCGGAAGTTAAGCGCGCGCGGCGGTACGCGACGCGCACCTGCGCGGCTTCAGCG	29014
Oy	496	GAGAGGACGACCGGACGAGAGCGCGGCGCACCGGCGCTTGACGACGCGTCTCTTC	555
Db	29013	CGCTCGGCGTCCGGGGCGGGGACAGGGGAGGCGGTGAGGGGTGAGTCCGTGCCG	28954
Oy	556	TTGCGGCTCGCCTCGCGCGCTCGCGCTGACGTCGCGGCGCTGCGCGGCGACGCGACG	615
Db	28953	GCCAGCGCGGACGAGGACCGGTGTCGGGCGGGGTGAGCGCGCTCGTCCGTCGCGCGG	28894
Oy	616	CCGCGCCCAACAGGTGCTGCTCGCGCTCTGCGCGCGGTGCTGCGCGGACGCCGAGGT	675
Db	28893	CCGCTTCAGCGCGGGGACGTAACCTGTCGCGCGCGGCGGACCGCGTCTCGGTTCCCG	28834
Oy	676	TACGGGACATTCAGCGGTCGACGCCCGCGGCGGCGTTCGCGCGCGCGCTCGCGAG	735
Db	28833	GCGAGGGCGGTGCTGCTAGCGCGCGCGGGGAGCGCGGAGTGTGTCGAGTTCACGAGCG	28774
Oy	736	GCGCTGCGCTCTTTCGCGCGGCTGCTGCGCGGACCGCGAGCGGCGCGCGCGG	795
Db	28773	GAAACGGGCGGGGCTCTCGCGCTCGCGCGGAGCGGACGAGCGGCCACCGCGCGCGG	28714
Oy	796	CGCGCGGACGCGCGCGGACATTCGCGGGGAGTGCAGCGGCGCGGAGACTTCGACGCGCT	855
Db	28713	CGGGTGTTCACCTCGGCGGTGTCGCGTACGCGACCGGCGCGCGGGGTGGCACCAACG	28654
Oy	856	TTGCGGGAATTCGCGCGGCTCATGGGACCGAGCTGCGGCGGGCGGAGCACTGTCG	915
Db	28653	GCGGCGCGCTCCGCGCGCTCGCGGCGCGGAGCGCGCGCTGACGAGGTTCAGACCG	28594
Oy	916	GAGCTGTTCCGACCTTCACCGGACCGACGAGCGAGG	953
Db	28593	GCAAGGCCAGGTGTGCGGTGCGCGGCGGACGTCGCGG	28556

```

RESULT 7
US-09-105-537-30
; Sequence 30, Application US/09105537A
; Patent No 6265202
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30
; LENGTH: 13842
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-30

```


Query Match 10.7%; Score 104; DB 3; Length 13842;
Best Local Similarity 45.1%; Pred. No. 3,7e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CGTCCGCGGAGCGGGGCGGTGGCGGCGATGTCGCGGCGCTGCTGGGGAGCGGCGAG 86
DB 12219 CGACGGACACATCAGGTGGGCTGGCGCGCTGGCTCCGTCGCGCGCGAGCCGCGAGCCGCGC 12278
QY 87 CCGCAACGCTGCTGTCGACCTGTCGTAACCGCGCGCGGAGCGGCGCGGACGCTGCTGGGG 146
DB 12279 ACCGGGCGCTGCGCCCGGAGGGGACGAGTCTGTCGACCGGCGGACCGCGGCTGGGGCG 12338
QY 147 CGAGTCAACGCGCGCGGCGCGGCGGAACTGCGCGCGCGCTTCGCGGAGCGGACCTGCTCT 206
DB 12339 ACTGATCGCGCGGACGATGTTGGGCGAGTGGGCGTACAGCGCTGCTGTTGAGCGC 12398
QY 207 GCTCCGCGTACAGAGGACGTCGCGCTCAAGGCGGTGGCGCGCGCTGACCGCGCTCATGCG 266
DB 12399 GCGGGGACGAGCGCGCGCGCGCGCGCGAGCTGTCGACGAGCTGAGGCGCTGGGAGC 12458
QY 267 GCGGCGCGCGCTGCTGCGCGACACCTGTCGTCGCGGACGCGGATGGCGCGGAGCTGCG 326
DB 12459 CGAGCTTCGCTGGGCGCGCTGCGACGTCGCGCGAGCGGAGCGCTCAACCGCGCTACGCA 12518
QY 327 GCGCCACGCGCGCGCGCGCTGCGACGTCGGGCTTCAACCGATGTTGCGCGCGCGCGCG 386
DB 12519 CGGCATCCCGCGGAAACCGCGCTCACCGCGCTGTCACACGCGAGCGCTCTCTCGCA 12578
QY 387 CATGACCGGCGGACCGCGCGCGCGCGGTGTCACAGGAGAGGCGCGCGCGCTCAACGCGCT 446
DB 12579 CGGCACTCTCCGCTCCATGACGAGGAGAGCTGGAACACGTACTGGGCGCGAAGGTCGA 12638
QY 447 GCTGCGGCTGTCGAGGCGCGCGCGCGCGCGCGCGCTGACGCGCTACGCGAGAGACGCA 506
DB 12639 CGCGCGCTTCTCTCTGACGAACTCACTGACGCGCGCGCTACGCGCTGCGACGCTTCTCT 12698
QY 507 CCGGACGACGCGCGCGCGACCGAGCGCTGACGACGCGGTGCTCTCTCTGCGGCTGCG 566
DB 12699 CATGTTCTCTCTCGCGCGCGCGCGCTTCTGCTGGGCGCGGCGCGCGCTACGCGCGCG 12758
QY 567 CCTCCGCGCGCTGCGGCTGACGTCGCGGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCA 626
DB 12759 CAAAGCCACCTTGGAGCGCTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 12818
QY 627 GGTGCTGCTGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
DB 12819 CCTCGGCTGGGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12878
QY 687 CCAAGCGTCCAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 746
DB 12879 CTGCGCGCGGATGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 12938
QY 747 CTTTCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
DB 12939 CCTGAGGCGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 12998
QY 807 CCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 864
DB 12999 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13058
QY 865 CTCGCGCGGCTCATGGGACGCGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 924
DB 13059 CGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13118
QY 925 CGCACTCTTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 966
DB 13119 CGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 13160

RESULT 8
US-09-105-537-5
; Sequence 5, Application US/09105537A
; Patent No. 6265202

; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600,438US1
; CURRENT APPLICATION NUMBER: US/09/105,537A
; CURRENT FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 5
; LENGTH: 36778
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-105-537-5

Query Match 10.7%; Score 104; DB 3; Length 36778;
Best Local Similarity 45.1%; Pred. No. 3,4e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CGTCCGCGGAGCGGGGCGGTGGCGGCGATGTCGCGGCGCTGCTGGGGAGCGGCGAG 86
DB 13960 CGACGGACACATCAGGTGGGCTGGCGCGCTGGCTCCGTCGCGCGCGAGCCGCGAGCCGCGC 14019
QY 87 CCGCAACGCTGCTGTCGACCTGTCGTAACCGCGCGCGGAGCGGCGCGGACGCTGCTGGTGG 146
DB 14020 ACCGGGCGTGGCGCCCGGAGGGGACGAGTCTGTCGACCGCGGACCGCGGCGCTGGGGCG 14079
QY 147 CGAGTCAACGCGCGCGGCGCGGCGGAACTGCGCGCGCGCTTCGCGGAGCGGACCTGCTCT 206
DB 14080 ACTGATCGCGCGGACGATGTTGGGCGAGTGGGCGTACAGCGCTGCTGAGCGC 14139
QY 207 GCTCCGCGTACAGAGGAGGTGGCGCTCAAGGCGGTGGCGCGCGCTGACCGCGCTCATGCG 266
DB 14140 GCGGGGACGAGCGCGCGCGCGCGCGCGAGCTGTCGACGAGCTGAGAGCGCTTGGAGC 14199
QY 267 GCGGCGCGCGCTGCTGCGCGACACCTGTCGTCGCGGACGCGGATGGCGCGAGCTGCG 326
DB 14200 CGAGCTTCGCTGGGCGCGCGTGGGAGGTCGCGCGAGCGGAGCGCTCAACCGCGCTACGCA 14259
QY 327 GCGCCACGCGCGCGCGCGCTGACGAGTGGGCTTCAACCGATGTTGCGCGCGCGCGCG 386
DB 14260 CGGCATCCCGCGGAAACCGCTTCAACCGGTGTCACAGGAGCGGTCTCTCGCA 14319
QY 387 CATGACCGGCGGACCGCGCGCGCGCGGTGTCACAGGAGCGGCGCGCGGTGTCACGCGCT 446
DB 14320 CGGCACTCTCCGCTCCATGACGAGGAGAGCTGGAACAAGTTCGCGCGCGCGCGCGCG 14379
QY 447 GCTGCGGCTGTCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 506
DB 14380 CGCGCGGCTTCTCTGTCGAGGAACTCACTGACGCGCGCGCGCGCGCGCGCGCGCGCG 14439
QY 507 CCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 566
DB 14440 CATGTTCTCTCTCGCGCGCGCGCGCTTCTGATGCGCGGCGCGAGCGCGCTTACGCGCG 14499
QY 567 CCTCCGCGCGCTGCGGTGAGAGTCCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 626
DB 14500 CAAAGCCACCTTGAAGCGCTTGCCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTC 14559
QY 627 GGTGCTGCTGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 686
DB 14560 CCTCGGCGTGGGCGCTTGGGCGGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 14619
QY 687 CCAAGCGTCCAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 746
DB 14620 CTTGCGCGCGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 14679
QY 747 CTTTCGCGCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 806
DB 14680 CCTTCAGCGCGCGCGCTTCCGCGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 14739

Accession	Sequence	Position
Oy	807 CCGGGGACATCCG--GGGATGGGAGGGGCGGGAACTGACGGGCGCTTTCGGGGAA	864
Db	14740 CGCCGGGCTGGGGAGCCGGGCGGGGAAAGACCCGGCCGGATCCGGCGCTTTCGGGGA	14799
Oy	865 CTCCGCGGCTCATGGGAGCCGGAGCTGTCGCGCGGGCCAGGACCACTGCCAGAGCTGTTTC	924
Db	14800 CGTGTGTGGGCCCAAGAACCTGCCGGGCCCGGCGGCTCCGGCTTCGACGACGACG	14859
Oy	925 CGCACCTTCACCGGACCGAGAGAGGCGAGAGAGACCGA	966
Db	14860 CGGGACGGCCCGGACGCGCCGGGAGCGGCGACGCGCGGCGGA	14901

RESULT

```

US-09-320-878-19
: Sequence 19, Application US/09320878A
: Patent No. 6117659
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary C.
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
: FILE REFERENCE: 30662200210
: CURRENT APPLICATION NUMBER: US/09/320, 878A
: CURRENT FILING DATE: 1999-05-27
: EARLIER APPLICATION NUMBER: CIP OF 09/141, 908
: EARLIER FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: 60/119, 139
: EARLIER FILING DATE: 1999-02-08
: EARLIER APPLICATION NUMBER: 60/100, 880
: EARLIER FILING DATE: 1998-09-22
: EARLIER APPLICATION NUMBER: 60/087, 080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 34
: SOFTWARE: Patentin Ver. 2.0
: SEQ ID NO 19
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-320-878-19

```

Query Match	10.7%	Score 104;	DB 3;	Length 38506;
Best Local Similarity	45.1%	Pred. No. 3.3e-08;		
Matches 425; Conservative	0;	Mismatches 515;	Indels 2;	Gaps 1;

Qy	27	GTGTGGGCGGACGGGGGCGGTGGGCGGCATGTTTCGCGGGCTCTGTCGGGAGCGGGCGACG	86
Db	12102	CGACGGACCACTACAGGCTGGGCGGCTCTGGCTCTGGTCGGGCCGGAACCGGACCGCCCGC	12161
Qy	87	CCGACGCTGTCGTCTGACCTCTGTACCGCGCGCGGACCGGCAGCGCTGCTCTGTGTGG	146
Db	12162	ACCGGCGGTGCGCCCGGAGGGACAGTCTGTGACCGCGGGACCGGGGGGCTGTGGCGG	12221
Qy	147	CGAGCTACCGCGCGCGGGGCGCCGAACTCGCGGCGCGCTTCTCGGAGACCGGACTCTGCTT	206
Db	12222	ACTGTCGCGCCGCGACCTGTGTGGGCGAGTGGGGGTACGACGCTCTCTGTGTGACCGG	12281
Qy	207	GCTGCGCGTACAGAGGACGTGGGCTTCAAGAGGCGTGGGCGCGGTACCGGCTCATGCG	266
Db	12282	GGGGGGACGGACCGCCCCGGGCGCCGACGAGCTGTGTACAGACTGGAGGCTTGTGGAGC	12341
Qy	267	GCCGGGCGCGCTGTGCTGCGGACACCTCTGTCCGTCCGACCGGCGATGAGCGCGAGCTTCG	326
Db	12342	CGAGCTCTGGGTGGCGGGGTGCGAGCTGCGCGACCGGAGAGCCCTACCGCGGTACTTCA	12401
Qy	327	GGCCACGCGCCCGGCGTCCAGCACGTGGGCGCTCAACCGATGTTGGCCCCCGCGCCCGG	386

Db	12402	GGCATTCCCGCGGAAACCCCGCTCACCGGAGTCTTCACACGGACGGGTCCTTCACGA	12461
Qy	387	CATACCGGCGCGAACCCGTGGCCGCCCGTGGTCAACAGAGAACGGGCGGGGCTCAACGGCCCT	446
Db	12462	CGGACCCCTCCCGTTCATGACGACGAGAGACGTGGAAACGTATCGCGGCCCAAGTTCGA	12521
Qy	447	GCTCGGCTGTCTGAGAGGCGCGGCGGACAGGCCCTGACGGCTTCAACGCGGAGAGACGA	506
Db	12522	CGCCGCGTTCCTCTCTGACGAACTACACTTCGACGCCCGCATAGACTTGGCAGCGTTGGT	12581
Qy	507	CCGGACGACGCGCGGCCACCCAGGCCCTGACGACCGCGTGGCTCTCTCTTGGGAGCTGGC	566
Db	12582	CATGTCTCTCCGCGCGCGCGCGCTGTCTTGGATGGCGGGGCGAGGGCGCTTACGCGCGCG	12641
Qy	567	CCTTGGCCCGCTTGGGGTGCAGCTCCGGGCGCTTGGGCGGACGAGCACCGCGCCCTCACCA	626
Db	12642	CAAGGCCACCTTCGACCGCCCTTCCTCGCGGGCGCGCGGGAGCGGACCTTCCGCGCTTTC	12701
Qy	627	GGTGTCTGTGCGCCTCCGTGCGCGGTGTGCTCGGCGGACGCCCGAGGTGTACGGGACAT	686
Db	12702	CCTTGGCTGGGAGCTTCTGGGCGCGAGACGACGGGCAATGACCGGCGAGCTCGGCGAGCGGA	12761
Qy	687	CCAGCGGTCCAACCCCCCGGCGGGCGTTCGCGCGCGCGGCGCTTTCGCGAGGACCTTGGCGTTC	746
Db	12762	CTTGGCGCGATGACGCGCGCGGACATCGGCGGATCAGCGACGCGGAGGACATCGCGCT	12821
Qy	747	CTTTCGCGCGCTGTGTGGGACGACACCCGGAATCGTTCGACACGCCCTCCGGGCGCGCGACGC	806
Db	12822	CTTCGAGGCGCGCCTTCGCGACACCGGACGCGGACGCGTGTCTGTGCGCCCTTGGCGCTGACGCG	12881
Qy	807	CCCGGCGCATTCGCG--GGGATGTCGACGGGCGCGGGAACTTGACGCGCTTCTTGGGGGAA	864
Db	12882	CGCCGGGCTGCGGAGACCGCGCGCGGAAACACCCGCGGGAATCCCGGCGCTCTTCCGGGA	12941
Qy	865	CTCCGCGCGTCAATGGGACCGGAGCTTCGCGGCGGAGCACGACCACTTGCAGGAGCTGTTC	924
Db	12942	CGTGTGTGGGCGACGAGACGCTTCGGGCGCGGCGCGTTCGCGGCGCTTCGCTTACGACAGC	13001
Qy	925	CGACACCTTCACCGGACCGGACGAAAGCGAGAGAGACGGA	966
Db	13002	CGGAGCGGCGGACGCGCGGGGACCGCGGAGACGCGCGGCGGA	13043

RESULT 10

```

US-09-141-908-1
: Sequence 1, Application US/09141908
: Patent No. 6503741
: GENERAL INFORMATION:
: APPLICANT: ASHLEY, Gary
: APPLICANT: BETLACH, Melanie C.
: APPLICANT: BETLACH, Mary
: APPLICANT: MCDANIEL, Robert
: APPLICANT: TANG, Li
: TITLE OF INVENTION: Combinatorial Polyketide Libraries Produced Using a
: TITLE OF INVENTION: Modular PKS Gene Cluster as Scaffold
: FILE REFERENCE: 300622002100
: CURRENT APPLICATION NUMBER: US/09/141,908
: CURRENT FILING DATE: 1998-08-28
: EARLIER APPLICATION NUMBER: CIP OF 09/073, 538
: EARLIER FILING DATE: 1998-05-06
: EARLIER APPLICATION NUMBER: CIP OF 08/846, 247
: EARLIER FILING DATE: 1997-04-30
: EARLIER APPLICATION NUMBER: PROV. 60/076,919
: EARLIER FILING DATE: 1998-03-05
: EARLIER APPLICATION NUMBER: PROV. 60/087,080
: EARLIER FILING DATE: 1998-05-28
: NUMBER OF SEQ ID NOS: 31
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1
: LENGTH: 38506
: TYPE: DNA
: ORGANISM: Streptomyces venezuelae
US-09-141-908-1

```

Query Match 10.7%; Score 104; DB 4; Length 38506;
Best Local Similarity 45.1%; Pred. No. 3.3e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CATTGCGGAGAGCGGGGCGGTGGCGGCAATGTTGCGGGGCTGCTGGCGGAGCGGGACG 86
DB 12102 CGACGGGACCATGAGGTGGCGCGGCTGGGCTCTGTCGGCGCCGAGACCGGACCGCGCG 12161
QY 87 CCGGACGCTGCTGCTGACCTTCGTAACCGCGCGGAGCGGGCGGAGCGGCGCTGGTGGG 146
DB 12162 ACCGGGCGCTGGCGCGGAGGAGGACGCTGCTGTAACCGGCGGAGCGGGCGGCGGCGG 12221
QY 147 CGAGTTCAGCGCGCGGGGCGGAACTGCGGCGCGGCTTCGGGAGCGGAGCTCTGCT 206
DB 12222 ACTGATGCGCGGAGCGTGGGCGAGTGGGGGTGACGACGCTGCTGCTGGTGAAGCG 12281
QY 207 GCTGCGGCTACAGAGGAGCTGCGCTCAAGGCGGCTGGCGCGGCTGACCGGCTCATGCG 266
DB 12282 GCGGGGCGACGAGACGCGCGGCGCGCGGACGAGCTGCTGACAGAGCTGGAGGCGCTGGGAGC 12341
QY 267 GCGGCGGCGGCTGCTGCGCGGACACCTGCTCGTCCGAGCGGGGATGGCGCGGAGCTGCG 326
DB 12342 CGACGTTTCGCTGGCGCGGCTGCGACGTCGCGACCGGAGACCTCAACGCGCTGACTCGA 12401
QY 327 GCGGCGGCGCGCGCGGCTGCGAGACGTTGGGCTCAACCGGATGTTGCGCGCGCGCGCG 386
DB 12402 CGGCACTCCGCGGAGAACCGCGGCTCACCGGGTGTGTCACACGCGAGCGCTCTCTCGA 12461
QY 387 CATGACCGGCGGACCGCTGGCGCGCTGTGTGTCACAGGAGATGGCGCGGCGCTCACGCGCT 446
DB 12462 CGGACACCTCCGCTCCATGACGAGAGGAGCTGTGACACGTTCTGGGCGCTCAAGGTGCA 12521
QY 447 GCTGCGGCTGTGAGAGGCGGCGGCGGAGCGCGCTGACGCGCTGACGCGCGGAGAGACGA 506
DB 12522 CGCGGCTTCTCTCTGACGAACTACCTGACGCGCGCTGACCTGGCAGGCTTCTGCT 12581
QY 507 CCGGACGAGCGGCGGACCGGAGCGCTGACGACGCGGCTGCTCTCTTTCGAGGCTGCG 566
DB 12582 CATGTTCTCTCCGCGCGCGCGCTTTCGTTGGGCGGGGAGAGGCGCTTACCGCGCGCG 12641
QY 567 CTTGCGCGGCTGCGGCTGCAAGCTCGGAGCGCTGAGCGGCGGCGGACGCGCGCGCTCA 626
DB 12642 CAACGCGACCTCTGACGCGCTGCTGCGCGCGCGCGCGGAGCGGAGCTCCCGCTCTC 12701
QY 627 GGTGCTGCTGCGCGCTGCGCGCGGCTGCTGCGGCGGAGCGCGGAGGTGTAAGGGGACAT 686
DB 12702 CTTGCGGCTGGAGGCTTCTGGGCGGAGACGAGCGGAGTATGACGAGGAGCTCGGCAAGGCGA 12761
QY 687 CCAAGCGGTCCACCCCGGCGGCGGCTGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCTG 746
DB 12762 CTGCGCGGAGTGAAGCGGCGGCGGAGTCCGCGGAGTACGAGGAGCGGAGGCACTCGGCT 12821
QY 747 CTTGCGCGGCTGCTGCGGAGCAAGCTCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806
DB 12822 CTTGCGAGCGGCGGCTCTGCGGAGCAAGCGGCGGCTGCTGCTGCTGCGGCTGCAAGCG 12881
QY 807 CCGCGGCGCATGCGG--GGGATGAGAGCGGCGGCGGAGACTGACGCGGCTCTTTCGAGGAG 864
DB 12882 GCGCGGCGCTGCGGAGCGGCGGAGAGCGAGCGGCGGAGTCCCGGCGCTTTCGAGGAG 12941
QY 865 CTCGCGGCGCTGATGAGAGCGGAGCTGCGGCGGCGGAGAGCACTGCGAGAGCTGTTCT 924
DB 12942 CGTGTGCGGCGGAGAGCGTCCGCGGCGGCGGCGGCTGCGGCTCTGCGAGAGAGC 13001
QY 925 CGGACCTTCCAGCGGACCGAGCGAGAGGCGGAGAGAGAGCGGAG 966
DB 13002 CGGAGCGGCGGAGCGGCGGAGAGCGGAGAGCGGCGGAGAG 13043

RESULT 11
US-09-657-440-19
; Sequence 19, Application US/09657440

/ Patent No. 6509455
/ GENERAL INFORMATION:
/ APPLICANT: ASHLEY, Gary
/ APPLICANT: BETTACH, Melanie C.
/ APPLICANT: BETTACH, Mary C.
/ APPLICANT: McDANIEL, Robert
/ APPLICANT: TANG, Li
/ TITLE OF INVENTION: RECOMBINANT NARBONOLIDE POLYKETIDE SYNTHASE
/ FILE REFERENCE: 30622002120
/ CURRENT APPLICATION NUMBER: US/09/657,440
/ PRIOR FILING DATE: 2000-09-07
/ PRIOR APPLICATION NUMBER: 09/320,878
/ PRIOR FILING DATE: 1999-05-27
/ PRIOR APPLICATION NUMBER: CIP OF 09/141,908
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 19
/ LENGTH: 38506
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-09-657-440-19

Query Match 10.7%; Score 104; DB 4; Length 38506;
Best Local Similarity 45.1%; Pred. No. 3.3e-08;
Matches 425; Conservative 0; Mismatches 515; Indels 2; Gaps 1;

QY 27 CATTGCGGAGAGCGGGGCGGTGGCGGCAATGTTGCGGGGCTGCTGGCGGAGCGGGACG 86
DB 12102 CGACGGGACCATGAGGTGGCGCGGCTGGGCTCTGTCGGCGCCGAGACCGGACCGCGCG 12161
QY 87 CCGGACGCTGCTGCTGACCTTCGTAACCGCGCGGAGCGGGCGGAGCGGCGCTGGTGGG 146
DB 12162 ACCGGGCGCTGGCGCGGAGGAGGACGCTGCTGTAACCGGCGGAGCGGGCGGCGGCGG 12221
QY 147 CGAGTTCAGCGCGCGGGGCGGAACTGCGGCGCGGCTTCGGGAGCGGAGCTCTGCT 206
DB 12222 ACTGATGCGCGGAGCGTGGGCGAGTGGGGGTGACGACGCTGCTGCTGGTGAAGCG 12281
QY 207 GCTGCGGCTACAGAGGAGCTGCGCTCAAGGCGGCTGGCGCGGCTGACCGGCTCATGCG 266
DB 12282 GCGGGGCGACGAGCGCGCGGCGCGGAGCTGTGTCACAGCTGAGAGGCGGCGGAGGAGC 12341
QY 267 GCGGCGGCGGCTGCTGCGGAGACGTTGGGCTCAAGGCGGCGGCGGCGGCGGCGGCTG 326
DB 12342 CGACGTTTCGCTGGCGCGGCGGCTGCGACGTCGTCGTCGACGAGCTGAGAGGCGGCGG 12401
QY 327 GCGGCGGCGGCGGCGGCTGCGAGACGTTGGGCTCAACCGGATGTTGCGCGCGCGCG 386
DB 12402 CGGCACTCCGCGGAGAACCGCGGCTCACCGGGTGTGTCACAGGAGGCTCTCTCGA 12461
QY 387 CATGACCGGCGGACCGCTGGCGCGGCTGTGTGTCACAGGAGATGGCGCGGCGCTCACGCGCT 446
DB 12462 CGGACACCTCTGCTCATGACGAGAGGAGCTGTGAACTGTATGCGGCGGCGGAGGTGCA 12521
QY 447 GCTGCGGCTGTGAGAGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 506
DB 12522 CGCGGCTTCTCTCTGAGAGAACTACCTGACGCGCGGATGACAGCTGAGAGCTTCTGCT 12581
QY 507 CCGGACGAGCGGCGGACCGGAGCGCTGACGAGCGGCTGCTCTCTTTCGAGGCTGCG 566
DB 12582 CATGTTCTCTGCGCGCGGCGGCTTTCGTTGGGCGGGGAGAGGCGGCTTACGCGCGCG 12641
QY 567 CTTGCGCGGCTGCTGCGGAGCAAGCTCGGAGCGGCGGCGGAGAGAGAGCGGCGGCGGCGG 626
DB 12642 CAACGCGACCTCTGAGCGGCTTGGCGGCGGCGGCGGAGCGGAGACTTCCCGCTCTC 12701
QY 627 GGTGCTGCTGCGGCTTCTGAGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 686
DB 12702 CTTGCGGCTGGAGGCTTCTGGGCGGAGAGCGGCGGAGTATGACGAGGAGCTCGGCAAGGCGA 12761
QY 687 CCAAGCGGTCCACCCCGGCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTGCGGCT 746

Db 12762 CTGCGCCGGATAGCCGCGCGGATCGGCGGGATTCAGGACCGGAGCATCGCGCT 12821
QY 747 CTTGCGCGCGCTGTCGCGGACGACCGGACCGGTCGCGACCGCCCGGCGCGCGGACGCG 806
Db 12822 CTTGCGGACCGCGCTCTCGGACGACCGGACCGGTCGTCGCTCGCTTGGGCTCGAGCG 12881
QY 807 CCGCGCGCATCCCG--GGGANTGCGAGCGCGCGGGAACCTCGACGCGCTCTTGGGGA 864
Db 12882 CCGCGGGCTGCGGAGCGCGCGCGGGAACGACCGCGCGGAAATCCCGCGCTCTTGGGGA 12941
QY 865 CTCGCGCGGCTCATGGGACCGGAGCTGCGCGCGGCGGACGACCATGCGAGCTGTC 924
Db 12942 CGTGTCGCGCGCGGACCGCTCCGCGGCGCGCGCTCGCGCTCGCGCTCGAGACAGC 13001
QY 925 CGCACCTTCACGCGACCGGACGAGGAGGAGAGGACCGA 966
Db 13002 CGGAGCGCGCGGACCGCGGAGCGGCGGAGCGCGCGGGA 13043

RESULT 12

US-10-237-551-193
; Sequence 193, Application US/10237551
; Patent No. 6821519
; GENERAL INFORMATION:
; APPLICANT: Day, Craig H.
; APPLICANT: Hosken, Nancy A.
; APPLICANT: Parsons, Joseph M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DIAGNOSIS AND
; FILE REFERENCE: 210121.538C3
; CURRENT FILING DATE: 2002-09-06
; NUMBER OF SEQ ID NOS: 254
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 193
; LENGTH: 3957
; TYPE: DNA
; ORGANISM: HSV2
US-10-237-551-193

Query Match 10.6%; Score 102.6; DB 4; Length 3957;
Best Local Similarity 46.8%; Pred. No. 6.8e-08;
Matches 439; Conservative 0; Mismatches 484; Indels 16; Gaps 3;

QY 7 GGGTTCCCGCGCGCTGTCGTCGCGCGGACCGGCGGTCGCGCATGTCGCGCG 66
Db 1695 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1752
QY 67 CTGCTCGGAGGCGGCGGACCGGACGCGCTGTCGACCTGTCACCGCGCGCGGAGCG 126
Db 1753 GCGGTCCCGCGCGGCTACGCGCGCGCGGAGGTCCTGCGCGCTGCGGCGCGTGAAGCGC 1812
QY 127 CCGGAGCGCTGCTGTCGTCGTCGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 186
Db 1813 GCGCGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 1872
QY 187 CCGGAGCGGACCTCGTCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 246
Db 1873 GGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1932
QY 247 CCGGTACCGCGCTCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 1933 CCGGAGATCTGAGAGCGCGCTGCGGAGGCTTTCGACGCGACCTGCGCGCGCGCGCGCG 1992
QY 307 GCGATGCGCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 366
Db 1993 CTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2052
QY 367 ATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db 2053 CCGGAGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2112
QY 427 GGGCGCGGCGTCAAGCGCGCTGTCGCGCGCTGTCGAGGAGGCGCGCGCGCGCGCGCGCG 486

Db 2113 GCGCTGTGCTGATGCGCTGCGCGGAGACCTGCGCTGAGCGCGCGCGCGCGCGCGCGCG 2172
QY 487 CTCACGCGGAGAGACGACCGGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Db 2173 GTGCGCGCGGTGCGCGCGCTGTGAGCGCTGTGTCGCGCGGCGCGCTGCGCGCGCG 2232
QY 547 CTCCTCTCTTGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
Db 2233 AGCCCGCGCTGTGAGCTTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2292
QY 602 CCGGAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
Db 2293 AGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2352
QY 662 GCAAGCGCGAGGTATGCGGAGCATTCAGCGCTCAACCGCGCGCGCGCGCGCGCGCGCG 721
Db 2353 TCGCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2412
QY 722 GGGCGCTGCG-----GAGGCGCTGCGCGCGCTTTCGCGCGCGCTGTCGCGGACGCG 772
Db 2413 GGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2472
QY 773 CCGACCGTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 832
Db 2473 CCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2532
QY 833 GCGCGCGGAACTTCGAGCGCGCTTTCGCGGAACTCCGCGCGCTCATGAGACGAGCTG 892
Db 2533 GCGCTTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2592
QY 893 CCGCGCGCGGACGACCTGCGGAGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931
Db 2593 CCGGCGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2631

RESULT 13

US-09-827-688-8
; Sequence 8, Application US/09827688
; Patent No. 6821955
; GENERAL INFORMATION:
; APPLICANT: ORSON, FRANK
; APPLICANT: KINSEY, BERNIA
; APPLICANT: BHOAGL, BALBIR
; TITLE OF INVENTION: MACROMERGED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION
; FILE REFERENCE: P01949US1/10004014
; CURRENT FILING DATE: 2001-04-06
; PRIOR APPLICATION NUMBER: 60/195,680
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8
; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

Query Match 10.6%; Score 102.6; DB 4; Length 154746;
Best Local Similarity 46.8%; Pred. No. 4.8e-08;
Matches 439; Conservative 0; Mismatches 484; Indels 16; Gaps 3;

QY 7 GGGTTCCCGCGAGCGTCTGTCGCGCGGACCGGCGGCGGTCGCGCATGTCGCGCG 66
Db 151404 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151461
QY 67 CTGCTCGGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 126
Db 151462 GCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151521
QY 127 CCGGAGCGCTGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 186

Db 151522 GCGCCCGGCTCCGGCCGGGCGGCGGCGGACGACACGACGACGACGACGACGCGGCGCGGCT 151581

Qy 187 CGGAGCGCGACCTTCCTCTCTGCGCGGTACACAGAGCTGGGCCCTTCAGAGCGCTGGCG 246

Db 151582 GGTGGCGCGCGCGCGCGCGCGGAGCGGCGCGCGCGCGCTGGAGTGGCTGGCGCGCTGC 151641

Qy 247 CCGGTGACCGCGCTCATGCGGCGCGGCGCGCGCTGGCGCGACCCCTGTCCGTCCGAGC 306

Db 151642 CGCGGGATCTCTGAGAGCGCTGAGCGGAGGCTTCGACGGCGACCTGGCGCGCTGCTCCGAG 151701

Qy 307 GGCATGCGCGCGGAGCTGCGGCGCGCGCGCGCGCGCGCGCTGCGACAGTGGGCTTCAACCG 366

Db 151702 CTGGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151761

Qy 367 ATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426

Db 151762 CGCGACCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 151821

Qy 427 GCGCGCGCGCGCTACCGCGCGCTGCGCGCTGCGCGCTGCGAGGCGCGCGCGCGCGCGCGCTACG 486

Db 151822 GCGCTGGTGTGATGATGCGCGCTGCGCGGAGACCTTGGCGCGTGGCGCGGAGGAGCGCGC 151881

Qy 487 CTCAACGCGGAGAGAGACGACCGCGACGACGCGCGCGCACCCAGCGCGCTTGAAGCAGCGCGTG 546

Db 151882 GTGGCGCGCGCTGCGCGCGCGCGCGGTAGCTGCTCGCGCGCGCGCGCGCGCGCGCGCG 151941

Qy 547 CTCTCTCTCTTCCGAGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601

Db 151942 AGCGCGCGCGCTGTGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152001

Qy 602 CGCGGACGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661

Db 152002 AGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152061

Qy 662 GCGCGCGCGAGGTGATCGGGGACATCCAGCGGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCG 721

Db 152062 TCCGCGCGCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 152121

Qy 722 GCGCGCGCTGC-----CGAGCGCTGCGCTCTTGCAGCGCGCTGGTCCGCGACGAC 772

Db 152122 GCGCGCGCGCGCGCGCGCGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 152181

Qy 773 CGGACCGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 832

Db 152182 CCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152241

Qy 833 GCGCGCGGAACTCGACGCGCGCTTTCGCGGGAATCCCGCGCGCTATGAGGACCGGACGCTCG 892

Db 152242 GCGCTGACCGCGCGCGCGCGCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152301

Qy 893 CGCGCGCGCGAGGACCTGCGCGAGGAGCTGTTCCGCGACCG 931

Db 152302 CGCGGCGCGACGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 152340

RESULT 14

US-09-827-688-8/c

Sequence 8, Application US/09827688

Patent No. 6821955

GENERAL INFORMATION:

APPLICANT: ORSON, FRANK

APPLICANT: KINSEY, BERMA

APPLICANT: BHOQAL, BALBIR

TITLE OF INVENTION: MACROAGGREGATED PROTEIN CONJUGATES AS ORAL GENETIC IMMUNIZATION P

TITLE OF INVENTION: AGENTS

FILE REFERENCE: P01949US1/10004014

CURRENT APPLICATION NUMBER: US/09/827,688

CURRENT FILING DATE: 2001-04-06

PRIOR APPLICATION NUMBER: 60/195,680

PRIOR FILING DATE: 2000-04-07

NUMBER OF SEQ ID NOS: 13

SOFTWARE: PatentIn version 3.0

SEQ ID NO 8

```

; LENGTH: 154746
; TYPE: DNA
; ORGANISM: HERPESVIRUS 2
US-09-827-688-8

```

Query Match	10.6%;	Score 102.6;	DB 4;	Length 154746;
Best Local Similarity	46.8%;	Pred. No. 4.8e-08;		
Matches 439;	Conservative	0;	Mismatches 484;	Indels 16; Gaps 3;

OY	7	363CTTCCCCCGAAGGTCGTGTCGAGGAGCAACGAGGACGAGTGAAGCGACATGTTTCGCCGGG	66
Db	130338	GCCCCGCGCCGCGCCGCCGCTTGACCTGTCGAGCGGCGCTGCGCGGC--CGACGAGCGC	130281
OY	67	CTGCTGCGAGGAGCGAGGACCGACGACTGTCGTGACCTGTACCGCGCCGCGAGACG	126
Db	130280	GCGGTGCCCCCGCTACGACGCGCCGAGGAGTGTCTGCGCCCTGTGGAGCGCTGACGCGC	130221
OY	127	CCGACGCGCTTGTCTGTGTGGCGACGTCACTGGCGCGGAGGCCGGAACCTGCGCGCGCTTC	186
Db	130220	GCGCCGCGCTTCGCGCCGCGCCGAGGCGCGACGACGACGACGACGACGACGCGCGCGGT	130161
OY	187	CGAGACCGGAGACTGTCTGTCGCGCCGACACGAGGAGCTGAGCCCTCAAGGCGGTGGCG	246
Db	130160	GATGCGCGCGGCGCGCGCGCGAGAGCGGCGCGCGCTGTGAGCTGTGAGTGTGCTGCGCTTCG	130101
OY	247	CCCGTGAACCGGCTCATGCGAGCGCGAGCGCGCTGTGCGACACCTGTCTCGTCCGAGCG	306
Db	130100	CGCGGGAATCTGTGAAGAGCGCTGGCGGAGGCTTGTAGAGGACCTGTGCGCGCGCGTGGCGGG	130041
OY	307	GGCATGCGCGCGAGACTGCGGCGCCACGCGCCCGCGGTTCAGACGCTGAGGCTTCAACCG	366
Db	130040	CTGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	129981
OY	367	ATGTTGCGCGCGCGCGCGCGCGCAATGACCGCGCGACCGCTGTGGCGCGTGTCAACAGAGAC	426
Db	129980	CCGACGCGCGAGCGCGCGCGCGCTTGCGCGCGCTGTGCTGTGCGAGCTGCGGTTTCGTGCGGAC	129921
OY	427	GAGCGGAGCGTTCACGAGCGCGCTGTGCGAGCTGTCTGTGAGAGCGAGCGCGAGCGCGCTTACG	486
Db	129920	GCGCGTGTGTGTATGCGCTGTGCGCGGAGACCTGTGCGCGGTATGAGCGCGGACGACGAGCGCGC	129861
OY	487	CTCACGCGCGAGAGAGACGACCTGGACGACGAGCGGCCACCTCAGAGCCTTGACGACGCGGTG	546
Db	129860	GTTGCGCGCGTGTGCGCGCGCTGTGAGCTGTGCGCGAGCGCGCTGTGAGCGCGCGCGCGCG	129801
OY	547	CTCCCTCTCTTCCGAGCTGCGCGCTGCGCGCGCGCTGTGAGAGTCCGAGCGCTG-----G	601
Db	129800	AGCGCGCGCTGTGAGCTCTGCGCGCGCGCGCGCGCGCGCGCGAGACTGTCTTTCAGAAACAG	129741
OY	602	CGGCGACGAGACCGCGCGCGCCACACAGTGTGCTGTGCGCTCTGAGCGCGGTGTGCTGTGAGCG	661
Db	129740	AGCGTGCGCCCTCTGTGTGTGCGACACCGTGTGCGCGCGCGAGACTGTGTGCCCGCGCGCGC	129681
OY	662	GCAGCGCGGAGGTGTACGAGGACATTCAGCGGTCCAACCCCGAGGCGAGGTCTCGCGGCGC	721
Db	129680	TCCGCGCGCGGAGAGCGCGCAAGGAGCCCGCGCCCGCGCGAGGCGCGCGCGCGGCG	129621
OY	722	GAGCGCTTGC-----CGAGGCGCTGTGCTCTTGTGCGCGCGCTGTGTGTGCGAGGAC	772
Db	129620	GCGCGCGCGCGCGCGCGCGCGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTT	129561
OY	773	CGAAGCGTGTGCGCGCGCGCGCGCGCGCGAGCGCGCGCGCGCGCATCCCGGAGGATGAGAG	832
Db	129560	CCCCCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	129501
OY	833	GCGCGGAGACTTCGACGCGCTCTTTCGAGGAACTCCGCGGCTCATGTGAGACCGGAGCTCTG	892
Db	129500	GCGGTGACCGCGCGCGCGCGCGAGGAGCGCGCGCGCGCGCGAGGAGGCGGTGTGCGCGGCGCG	129441
OY	893	CGGCGGCGCAAGACCATCTGCGAGGAGCTGTTTCGACCC	931
Db	129440	CCGCGGCGCGACGACGCGCGCGCGCTTGTGCGCGCGCGC	129402

```
RESULT 15
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patient No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match      10.4%; Score 100.4; DB 3; Length 4411529;
Best Local Similarity 45.3%; Pred. No. 7.7e-08;
Matches 405; Conservative 0; Mismatches 486; Indels 3; Gaps 1;

QY      2  TGAGCGGCTTCCCGGAGCGTCTGTCGCGGACGCGGCGCGGCGGCGCATTTTCG 61
DB      3934383  TGGCGGCTTGGCGCGCGGACACACCTTGGCGCGCGTTGAAATTGTTGGCGGCGCGCG 3934324

QY      62  CCGGGGCTGTCGGGAGGCGGCGGACCGCGACGCTGTCGTGACCTGTACCGCGCGG 121
DB      3934323  CCAAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTG 3934264

QY      122  GACGCGCGGACGCGCTGCTGTGGGCGGACGTACCGCGCGCGCGCGCGCGCGCGCGGAACTGCGCGCG 181
DB      3934263  GCTGTACGCGCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 3934204

QY      182  CCCTCGGAGCGGAGACCTGCTGCTGCGCGCGGTACAGAGAGCTGAGCCCTCAAGGCG 241
DB      3934203  CCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCG 3934144

QY      242  TGGCGCGCGGTGACCGCGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCC 301
DB      3934143  CCGCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3934084

QY      302  GAGCGGCGCATGCGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 361
DB      3934083  GCGCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 3934024

QY      362  ACCCGATGTTGCGCGCGCGCGCGCGCGCATGACCGCGCGACCGCGCGCGCGCGCGCGTGCACCA 421
DB      3934023  GCGCGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTACCGCG 3933964

QY      422  GGGACGGGCGCGCGCGTCAAGCGCGCTGCTGCGCGCTGCGAGGCGCGCGCGCGCGCGCG 481
DB      3933963  GCGCGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933904

QY      482  TACGGCTCACGCGCGGAGAGACGACCGCGCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 541
DB      3933903  CCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933844

QY      542  CCGTGTCTCTCTCTTTCGAGGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 601
DB      3933843  CCGTGTCTCTCTCTTTCGAGGCTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933787

QY      602  CCGCGACGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 661
DB      3933786  CCGCGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933727

QY      662  GCAGCGCGCGAGGTACGCGGAGCATCCAGCGGTCAACCGCGCGCGCGCGCGCGCGCGCGCG 721
```

```
DB      3933726  GCGCGCGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933667

QY      722  GGGCGCTGCGCGGAGCGCGCGCTGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 781
DB      3933666  GCGTGTGCTGTACAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933607

QY      782  CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 841
DB      3933606  CCGTACCGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933547

QY      842  ACCTCGACGCGCGTCTTTCGCGGAACTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 895
DB      3933546  CCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 3933493
```

Search completed: October 5, 2005, 10:06:57
Job time : 195.767 sec

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:50:02 : Search time 951.313 Seconds
(without alignments)
7033.027 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969

Sequence: 1 atgagcgcgtctcccccag.....aagcgagaagacgatga 969

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 742561 seqs, 345238358 residues

Total number of hits satisfying chosen parameters: 14865122

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:*

1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
20: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
21: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
22: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
23: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
24: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
25: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
26: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	969	100.0	969	US-10-472-587-5	Sequence 5, Appli
2	915.4	94.5	3305	US-10-126-927-68	Sequence 68, Appli
3	913.8	12391	14	US-10-126-927-67	Sequence 67, Appli
4	115.8	12.0	138203	US-10-819-386A-1	Sequence 1, Appli
5	115.6	11.9	125401	US-10-203-295-35	Sequence 35, Appli
6	114.8	11.8	65140	US-10-203-295-1	Sequence 1, Appli
7	109.2	11.3	4569	US-10-156-761-7339	Sequence 7339, Ap

Result No.	Score	Query Match	Length	ID	Description
8	109.2	11.3	9025608	US-10-156-761-1	Sequence 1, Appli
9	108.6	11.2	1884	US-10-156-761-3440	Sequence 3440, Ap
10	107.4	11.1	969	US-10-472-587-5	Sequence 5, Appli
11	107	11.0	1707	US-10-156-761-4723	Sequence 4723, Ap
12	107	11.0	9025608	US-10-156-761-1	Sequence 1, Appli
13	106.2	11.0	3305	US-10-126-927-68	Sequence 68, Appli
14	105.4	10.9	12391	US-10-126-927-67	Sequence 67, Appli
15	105.2	10.9	1167	US-10-156-761-6049	Sequence 6049, Ap
16	104	10.7	13842	US-09-861-289-30	Sequence 30, Appli
17	104	10.7	13842	US-09-860-846-30	Sequence 30, Appli
18	104	10.7	13842	US-09-988-384B-30	Sequence 30, Appli
19	104	10.7	13842	US-09-836-821-30	Sequence 30, Appli
20	104	10.7	13842	US-10-271-889-30	Sequence 30, Appli
21	104	10.7	36778	US-09-861-289-5	Sequence 5, Appli
22	104	10.7	36778	US-09-860-846-5	Sequence 5, Appli
23	104	10.7	36778	US-09-836-821-5	Sequence 5, Appli
24	104	10.7	36778	US-10-271-889-48	Sequence 48, Appli
25	104	10.7	37948	US-09-988-384B-5	Sequence 5, Appli
26	104	10.7	38506	US-09-793-708-19	Sequence 19, Appli
27	104	10.7	38506	US-10-201-365-1	Sequence 1, Appli
28	104	10.7	38506	US-10-160-539-19	Sequence 19, Appli
29	104	10.7	38506	US-10-468-828-19	Sequence 19, Appli
30	102.6	10.6	3957	US-10-200-562-193	Sequence 193, Appli
31	102.6	10.6	3957	US-10-237-551-193	Sequence 193, App
32	102.6	10.6	3957	US-10-945-050-193	Sequence 193, App
33	102.6	10.6	154746	US-09-827-688-8	Sequence 8, Appli
34	102.6	10.6	154746	US-09-827-688-8	Sequence 8, Appli
35	101.8	10.5	3024	US-10-939-036-35	Sequence 35, Appli
36	101.4	10.5	1839	US-09-758-759-150	Sequence 150, App
37	101.4	10.5	5802	US-10-152-886-84	Sequence 84, Appli
38	101.4	10.5	5802	US-11-053-576-84	Sequence 84, Appli
39	101.4	10.5	5802	US-11-053-055-84	Sequence 84, Appli
40	101.4	10.5	15738	US-10-329-079-46	Sequence 46, Appli
41	101.4	10.5	61944	US-10-329-079-34	Sequence 34, Appli
42	101.4	10.5	109519	US-09-758-759-1	Sequence 1, Appli
43	100.4	10.4	2700	US-10-203-295-29	Sequence 29, Appli
44	100	10.3	88421	US-09-976-059-1	Sequence 1, Appli
45	99	10.2	2940	US-10-156-761-3788	Sequence 3788, Ap

ALIGNMENTS

RESULT 1

US-10-472-587-5

Sequence 5, Application US/10472587

Publication No. US20040214274A1

GENERAL INFORMATION:

APPLICANT: YANAI, Koji

APPLICANT: SUMIDA, Naomi

APPLICANT: WATANABE, Manabu

APPLICANT: MORIYA, Taisei

APPLICANT: MORAKAMI, Takechi

TITLE OF INVENTION: Transformants Producing Substance P/1022 Derivatives, Methods for

TITLE OR INVENTION: Producing The Same And Novel Biosynthesis Genes

FILE REFERENCE: 2003-1302A/MC/00144

CURRENT APPLICATION NUMBER: US/10/472,587

CURRENT FILING DATE: 2003-09-22

PRIOR APPLICATION NUMBER: 82227/2001

PRIOR FILING DATE: 2001-03-22

NUMBER OF SEQ ID NOS: 40

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 5

LENGTH: 969

TYPE: DNA

ORGANISM: Streptomyces venezuelae

FEATURE:

NAME/KEY: CDS

LOCATION: (1)..(966)

US-10-472-587-5

Query Match 100.0% Score 969; DB 20; Length 969;

Best Local Similarity 100.0%; Pred. No. 3.4e-192;

```
Matches 969; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGAGCGGCTTCCCGGAGCGTCTGTCGCGGAGCGGCGGATGGCGCATGTC 60
Db 1 ATGAGCGGCTTCCCGGAGCGTCTGTCGCGGAGCGGCGGATGGCGCATGTC 60
QY 61 GCCGGGCTGTCGGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 120
Db 61 GCCGGGCTGTCGGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 120
QY 121 GAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 180
Db 121 GAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 180
QY 181 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 240
Db 181 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 240
QY 241 GTGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 300
Db 241 GTGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 300
QY 301 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 360
Db 301 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 360
QY 361 AACCCGATGTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 420
Db 361 AACCCGATGTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 420
QY 421 AGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 480
Db 421 AGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 480
QY 481 GTAGCGGCTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 540
Db 481 GTAGCGGCTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 540
QY 541 GCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
Db 541 GCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
QY 601 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 660
Db 601 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 660
QY 661 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 720
Db 661 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 720
QY 721 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Db 721 CGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
QY 781 GCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 840
Db 781 GCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 840
QY 841 AACCTGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 900
Db 841 AACCTGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 900
QY 901 CAGGACCATGCTGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 960
Db 901 CAGGACCATGCTGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 960
QY 961 GACCGATGA 969
Db 961 GACCGATGA 969
```

RESULT 2
US-10-126-927-68

```
/ Sequence 68, Application US/10126927
/ Publication No. US20030082575A1
/ GENERAL INFORMATION:
/ APPLICANT: The Scripps Research Institute
/ APPLICANT: Schultz, Peter G
/ APPLICANT: Wang, Lei
/ APPLICANT: Anderson, John C
/ APPLICANT: Chin, Jason
/ APPLICANT: Liu, David R
/ APPLICANT: Magliery, Thomas
/ APPLICANT: Meggers, Eric L
/ APPLICANT: Mehl, Ryan A
/ APPLICANT: Pasternak, Miro
/ APPLICANT: Santoro, Stephen W
/ TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
/ FILE REFERENCE: 54-00120US
/ CURRENT FILING DATE: 2002-04-19
/ PRIOR FILING DATE: 2001-04-19
/ PRIOR APPLICATION NUMBER: US 60/285,030
/ PRIOR FILING DATE: 2002-02-06
/ NUMBER OF SEQ ID NOS: 79
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 68
/ LENGTH: 3305
/ TYPE: DNA
/ ORGANISM: Streptomyces venezuelae
US-10-126-927-68
```

Query Match 94.5%; Score 915.4; DB 14; Length 3305;
Best Local Similarity 96.6%; Pred. No. 38-181;
Matches 953; Conservative 0; Mismatches 16; Indels 18; Gaps 1;

```
QY 1 ATGAGCGGCTTCCCGGAGCGTCTGTCGCGGAGCGGCGGATGGCGCATGTC 60
Db 2007 ATGAGCGGCTTCCCGGAGCGTCTGTCGCGGAGCGGAGCGGATGGCGCATGTC 2066
QY 61 GCCGGGCTGTCGGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 120
Db 2067 GCCGGGCTGTCGGGAGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2126
QY 121 GAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 180
Db 2127 GAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2186
QY 181 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 240
Db 2187 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2246
QY 241 GTGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 300
Db 2247 GTGGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 2306
QY 301 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 360
Db 2307 CGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2366
QY 361 AACCCGATGTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 420
Db 2367 AACCCGATGTTCCCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 2426
QY 421 AGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 480
Db 2427 AGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2486
QY 481 GTAGCGGCTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 540
Db 2487 GTAGCGGCTACCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 2546
QY 541 GCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
```

[illegible]

RESULT 3

```

? Sequence 67, Application US/10126927
? Publication No. US20030082575A1
? GENERAL INFORMATION:
? APPLICANT: The Scripps Research Institute
? APPLICANT: Schultz, Peter G
? APPLICANT: Wang, Lei
? APPLICANT: Anderson, John C
? APPLICANT: Chin, Jason
? APPLICANT: Liu, David R
? APPLICANT: Magliery, Thomas
? APPLICANT: Meggers, Eric L
? APPLICANT: Mehl, Ryan A
? APPLICANT: Pastinak, Miro
? APPLICANT: Santoro, Stephen W
? APPLICANT: Zhang, Zhiwen
? TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
? FILE REFERENCE: 54-000120US
? CURRENT APPLICATION NUMBER: US/10/126,927
? CURRENT FILING DATE: 2002-04-19
? PRIOR APPLICATION NUMBER: US 60/285,030
? PRIOR FILING DATE: 2001-04-19
? PRIOR APPLICATION NUMBER: US 60/355,514
? PRIOR FILING DATE: 2002-02-06
? NUMBER OF SEQ ID NOS: 79
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 67
? LENGTH: 12391
? TYPE: DNA
? ORGANISM: Plasmid pSC101, Streptomyces venezuelae pAPABC
? US-10-126-927-67

```

```

Query Match 94.3% Score 913.8: DB 14 Length 12391;
Best Local Similarity 95.5% Pred. No. 4,3e-181;
Matches 952; Conservative 0; Mismatches 17; Indels 18; Gaps 1;

OY 1 ATGAGCGGCTTCCCCCGACAGCTGTCCTCGGCGGACACGGGGCGCGTGGCGGCATGTC 60
    |||||
Db 2361 ATGATGGCTTCCCCCGAGGCTGTCCTCGGCGGACAGGAGCGCGGCGGCATGTC 2420
    |||||

OY 61 GCCGGGCTGTCGGGAGCGGAGCGGACGCTCGTCGATCTTGATCGCCGCGC 120

```

Db	2421	GCCTGGCTGCTCGGAGAGGCGGAGCCGACCGTCTGTCGTCCGACCTTCGACCGCCG	2480
Oy	121	GSAGGCGCGAAGCGCTTACGTGTGGGAGGTCAACGCGCGGAGGCCGAACTTCGCGCC	180
Db	2481	GAAGGCGCGAAGCGCTCTCTGTGGGAGGTCAACGCGCGGAGGCCGAACTTCGCGCC	2540
Oy	181	GCCCTCCGGAGCGGAGCTTCGTCTGCTGCGGTACAGAGACGTGGCCCTTCAGGCC	240
Db	2541	GCCCTCCGGAGCGGAGCTTCGTCTGCTGCGGTACAGAGACGTGGCCCTTCAGGCC	2600
Oy	241	GTGGCGCCGTGACCCCGCTTCATGCGGCGCGGAGCGCGTGTGCGCCGACCCCTGTCCGTC	300
Db	2601	GTGGCGCCGTGACCCCGCTTCATGCGGCGCGGAGCGCGTGTGCGCCGACCCCTGTCCGTC	2660
Oy	301	CGAGCGGACATGCGCGGAGAGCTTCGGGAGCCACGCGCCCGGCGGTCCAGACGTGGGCTC	360
Db	2661	CGAGCGGACATGCGCGGAGAGCTTCGGGAGCCACGCGCCCGGCGGTCCAGACGTGGGCTC	2720
Oy	361	AACCCGATGTTTCGCCGCCCGCGCGCGGATGACCGGCGGACCCGTGGCGCGCGGTGTCAC	420
Db	2721	AACCCGATGTTTCGCCGCCCGCGCGCGGATGACCGGCGGACCCGTGGCGCGCGGTGTCAC	2780
Oy	421	AGGAGCGAGCGCGGAGCTCACGAGCTTCGTGCGGTCTGAGGAGCGGCGGCGGAGGCC	480
Db	2781	AGGAGCGAGCGCGGAGCTCACGAGCTTCGTGCGGTCTGAGGAGCGGCGGCGGAGGCC	2840
Oy	481	GTAAGGCTCAACGCGGAGAGACACGACCGGAGCCAGCGCGGCCACCCAGGCGCTTCAGCAC	540
Db	2841	GTAAGGCTCAACGCGGAGAGACACGACCGGAGCCAGCGCGGCCACCCAGGCGCTTCAGCAC	2900
Oy	541	GCCGTGCTCCTCTCTCCGAGGCTCGGCTTGCCCGGCTCGAGCGTCCGACTTCGAGGCTG	600
Db	2901	GCCGTGATCTCTCTCTCCGAGGCTCGGCTTGCCCGGCTCGAGCGTCCGAGGCTG	2960
Oy	601	GCGGCGACCGGACCGCGCGGCCACCAAGTGTGTGCTGCGCTTCGTGCGCGGTGTCTCGAC	660
Db	2961	GCGGCGACCGGACCGCGCGGCCACCAAGTGTGTGCTGCGCTTCGTGCGCGGTGTCTCGAC	3020
Oy	661	GGCAGCGCCGAGGTGACGGGAGACATCAGAGGTCCAAACCCCGGAGCGGCGTCCGCGGC	720
Db	3021	GGCAGCGCCGAGGTGACGGGAGACATCAGAGGTCCAAACCCCGGAGCGGCGTCCGCGGC	3080
Oy	721	CGGAGCGTCCGCGGAGGCTTCGCGCTCTTTGCGCGCTGGTCCGCGAC-----768	
Db	3081	CGGAGCGTCCGCGGAGGCTTCGCGCTCTTTGCGCGCTGGTCCGCGAC-----768	3140
Oy	769	-----GACCCGAGCCGTGCGGACGCGCCCGGGGCGCGGACGCGCCCGGACATCCCGGG	822
Db	3141	GCCGAGGACCCCGGACCGCGCGGACGACCCCGGACCGGACGACAAACCCCGGACATCCCGGG	3200
Oy	823	GGATGCGACGCGCGCGGAACTTCGACGCGGTCTTCGGGAACTTCGCGCGGCTTCATGGGA	882
Db	3201	GGATGCGACGCGCGCGGAACTTCGACGCGGTCTTCGAGGAACTTCGCGCGGCTTCATGGGA	3260
Oy	883	CCGAGGCTCGCGCGGCGCGGACCACTGCGAGAGCTGTTCCGACCTTCACCCGACCC	942
Db	3261	CCGAGGCTCGCGCGGCGCGGACCACTGCGAGAGCTGTTCCGACCTTCACCCGACCC	3320
Oy	943	GACGACGGAAGCGGAAAGACCGATGA	969
Db	3321	GACGACGGAAGCGGAAAGACCGATGA	3347

RESULT 4

US-10-819-386A-1/c
Sequence 1, Application US/10819386A
Publication No. US20050089884A1
GENERAL INFORMATION:
APPLICANT: Korea advanced Institute of Science and Technology
APPLICANT: Shanghai jiaotong University
APPLICANT: LEE, Sang Yup
APPLICANT: DENG, ZIXIN

```

; APPLICANT: CHEN, SHI
; APPLICANT: JEONG, Ki Jun
; APPLICANT: ZHOU, XIUFEN
; TITLE OF INVENTION: Genes for the Synthesis of FR-008 Polyketides
; FILE REFERENCE: P003-B015
; CURRENT APPLICATION NUMBER: US/10/819,386A
; PRIOR FILING DATE: 2004-04-06
; PRIOR APPLICATION NUMBER: KR10-2003-0074035
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 1
; LENGTH: 138203
; TYPE: DNA
; ORGANISM: Streptomyces sp. FR-008
US-10-819-386A-1
```

Query Match 12.0%; Score 115.8; DB 21; Length 138203;

Best Local Similarity 46.1%; Pred. No. 5.1e-16; Mismatches 497; Indels 12; Gaps 1;

Matches 436; Conservative 0; Mismatches 497; Indels 12; Gaps 1;

```

QY 6 CGGCTTCCCCGAGCGTGTGCTGCGGCGGAGCGGAGCGGAGCGGATGTTGCGCGG 65
DB 132882 CGGCCCCCGCTTCCGGCGCTGCGCGCGCTGTGGCCCGGGGCGAGAGAACCTTCCGCGA 132823
QY 66 GCTGCTCGGAGAGCGGCGAGCGGACGCTGTGTGACCTTGACCGCGCGGAGAGC 125
DB 132822 GGTGGCTCTGACGCGGAGACCGCGGGGCGACGCGCTCTTGTTGGCTGACCGCGCGCT 132763
QY 126 GCGGACGCGCTGCTGTGGGCGACGTACCGCGCGCGCGCGCGAAGTCCGCGCGCT 185
DB 132762 GCTGGAGCGGCGCTTCAAGCCACCGCTTCCCGCGCTTGGAGAGAGACCGCGCGCG 132703
QY 186 CGGAGACGCGGACCTGTCTGCTGCGCGGACGAGAGAGTGGCCCTCAAGGCGGTGC 245
DB 132702 CCGCGCGCTTGTGCTGGAGAGAGTCACTTCCACGCGCGCGCGCGCGAGGCGCGGT 132643
QY 246 GCGCGTGAACCGGCTGATGAGCGCGCGCGCTGTGCGCGACACCTGTCCGTCCGAGC 305
DB 132642 CCGCATACAC-----CCGCGGGGCGAGAGCGCGGTCACTTCACTGCGCGAG 132595
QY 306 GGGCATGCGCGCGGAGCTGCGGCGCGACCGCGCGCGCGCTGACAGAGTGGGCTCAAGC 365
DB 132594 CACGACCGGAGGACCGCGTGGCTGAGTGGCTCTGTGTCGCTTCCGCGCGCGAGCG 132535
QY 366 GATGTTGCGCGCGCGCGCGCGAGTACCGGCGCGAGCGCGCGCGCGGTGACAGGGA 425
DB 132534 GCGCGCGCGCGCGCGCGCGCGCGCTGTGGAGCGCGAGCGCGCTTCACTGAGAGCCG 132475
QY 426 CCGGCGCGGCGTACGCGCGCTGTGCGCGCTGTGAGAGGCGCGCGCGCGCGGTAGC 485
DB 132474 CTTCCACGCGGCGCGAGCGCGCGCGGAAACCGCGCGACCGTGGTCTGCGCGCGAGC 132415
QY 486 GCTCAAGGCGGAGAGAGACGAGACCGAGAGAGCGCGCGACCGCGCGCTTGAAGCGCGCT 545
DB 132414 CTTCCGCGTGTGGGCGCGCGCTTCAAGCGCGCGCGGAGTCAAGTGGCGCGTCAAGCG 132355
QY 546 GCTCTCTCTCTTGGGCTGCGCGCGCGCTTGGAGTGAAGTGAAGTGGCGCGCGCG 605
DB 132354 CCGCGGACCGAGACCGCGCGCGCTTGTGTCTTCCGCTGAGCGCGCGAGCGCGAGG 132295
QY 606 GACGAGCGCGCGCGCGCGCGAGTGTGTGCTGCGCGCTTGGCGCGGTGTGTGCGCGAG 665
DB 132294 CGTACCGCGAGGTGTCAGCGCGAGCACCGCGCGCGTCTTGAGCACATCAAGGTTGGCT 132235
QY 666 CCGGAGGTGTGTGAGGAGCATCCAGCGGTCAACCGCGCGCGCGCTCGCGCGCGCG 725
DB 132234 CCGCGGAGACACCGCGCGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 132175
QY 726 GCTGCGCGAGGCGCGCTGCTCTTGTGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 785
DB 132174 CCGCGAGGAGCGCTTCAAGACCGCGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGT 132115
```

```

QY 786 CGCCCCCGGCGCGCGAGCGCCCGCGCATCCCGGGAGATGCAAGCGCGCGGAACCT 845
DB 132114 CAGAGCGGAACACCGCCACCGCTTGGCTGTGACTGTGACCGCGGTGCGCGGAGACCA 132055
QY 846 CGAGCGCGCTTGTGGGAACTCGCGCGGTCAATGGAGCGGAGCTGTGGCGCGCGAGCA 905
DB 132054 CCGCGCGCGCTGTGCGCGCGCTTGGCGCGAGAGACCGCGCGCTGTGGCGCGCA 131995
QY 906 CCACTGCGAGAGCTGTTCGCAACCTTCCAGCGACCGAGAGCA 950
DB 131994 CACCGTGTTCGCGCGCGCGCGCTGCGCGCGCTGCGCGCGCGAGCGAGCA 131950
```

RESULT 5

US-10-203-295-35

; Sequence 35; Application US/10203295

; Publication No. US20040115762A1

; GENERAL INFORMATION:

; APPLICANT: Zolchev, Sergey Borisovich

; APPLICANT: Sekurova, Olga Nikolayivna

; APPLICANT: Fjaervik, Epsen

; APPLICANT: Brautaset, Trygve

; APPLICANT: Strom, Arne Reidar

; APPLICANT: Valla, Svein

; APPLICANT: Ellingsen, Trond Erling

; APPLICANT: Sletta, Havard

; APPLICANT: Gulliksen, Ole-Martin

; TITLE OF INVENTION: Novel genes encoding a nystatin polyketide synthase and their

; FILE REFERENCE: 1181-265

; CURRENT APPLICATION NUMBER: US/10/203,295

; PRIOR FILING DATE: 2003-05-19

; PRIOR APPLICATION NUMBER: PCT/GB 01/00509

; PRIOR FILING DATE: 2001-02-08

; PRIOR APPLICATION NUMBER: GB 0002840.7

; PRIOR FILING DATE: 2000-02-08

; PRIOR APPLICATION NUMBER: GB 0008786.6

; PRIOR FILING DATE: 2000-04-10

; PRIOR APPLICATION NUMBER: GB 0009387.2

; PRIOR FILING DATE: 2000-04-14

; NUMBER OF SEQ ID NOS: 49

; SOFTWARE: PatentIn version 3.0

; SEQ ID NO 35

; LENGTH: 125401

; TYPE: DNA

; ORGANISM: Streptomyces noursei ATCC 11455

US-10-203-295-35

Query Match 11.9%; Score 115.6; DB 19; Length 125401;

Best Local Similarity 48.5%; Pred. No. 5.8e-16; Mismatches 459; Indels 24; Gaps 4;

Matches 455; Conservative 0; Mismatches 459; Indels 24; Gaps 4;

```

QY 26 TCGTGGCGGCGAGCGGCGGTGTGGCGGATTTGCCCGCGCTGCTCGGAGCGCGGGA 85
DB 28560 TCGCGCGCGCTGTCCAGAGGCGTCCCGCGAGCGGCGCGCGCGCGCTTCTTCCGCA 28619
QY 86 GCGGACGCTGCTGTGCACTGTACCGCGCGCGGAGAGCGCGGAGCGCGCTGTGTGG 145
DB 28620 CCTTGAAGAGCGCGCGGTGTGACCGCGCGCGCTTGTGTGTGTGTGTGTGTGTGTGT 28679
QY 146 GCGAGTCAACGCGCGCGCGCGGACCTGCGCGCGCGCGCTTCCGAGAGCGGAGC--TGC 202
DB 28680 TCGCGGAGGAGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTG 28739
QY 203 TCTGTGCGCGGTACAGAGAGCTGTGCGCGCTTCAAGCGGTGTGTGTGTGTGTGTGTGT 262
DB 28740 CCGCGGTGAGCACCGCGAGCGCGCTTGTGCGCGCGCGCGCGCGCGCGCGCGCGAGC 28799
QY 263 TCGGCGCGGCGCGCGCTGTGCGGACCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCG 322
DB 28800 CCGAGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 28856
QY 323 TCGGAGCCACCGCGCGCGCGGTGAGCAAGTGGGCTTCAACCGATGTTGCGCGCGCG 382
```

```

Db      28857  TCGCGGTCGCGCGCACCGGCGTCCAGCCCGTGCCTCGTCCGTAAGCCGCGCGCGCG
Qy      383    CCGGCAATGACCGGCGGACCGCTGCGCGCGCTGCTACACAGGAGACGGGCGCGGCTACGG
Db      28917  ACGGTCCACACTGCTGCTGCGCACCGGACCGTCTGTTGTGGCGGACCGGCGGACCG
Qy      443    CCGTGTG-----CGGCTGTGACGGGCGGGCGGACCGGACCGGTAAGG
Db      28977  GCACCATGGCGCGCGCGCGCGCGCTGCGTGTGCGGAGGCGCGCACCTGCTC
Qy      488    TCAAGCGGAGAGACGACGACGAGACGAGCGCGGCAACCGGCGCTGACGCGCTGCG
Db      29037  TGACCGCGCGCGGACGAGACGACGCGCGGAGACGAGGCGCTGACG---GCGGAC
Qy      548    TCTCTCTCTTGGGAGTCTGCGCTGCGCGCTGCGCGCTGACGTCGGGCGCTGCGGCGA
Db      29094  TGGCGCGCTGCGCGCGCGGATGACCGTCTGAGACGACCGGACCGCGCGGACGCT
Qy      608    CGGACCGCGCGCGCGGACGAGTGTGCTGCGCTGCGCGCGGCTGCTGCGGCGGACG
Db      29154  TCGCGGCTCTCTGAGGAGTGTGCGGACGACCGCGCTGACCGCGGCTGCTGAGCGC
Qy      668    CCGAGTGTACGAGGAGATCAGCGGTCAACCCCGGCGGCGTCCGCGCGCGCGCGC
Db      29214  CGAGGCGGACGCGCGCGCGCGGACCGCGGCGGAGTGTGCGCGACGCGCGCGCTCA
Qy      728    TCGCCGAGGCGCTGCGCTCTTCTGCGCGCGGTGTGCGGACGACCGGACGTCGCG
Db      29274  CGCGCTTACGCGCGCGCTGCGCGCGCGCGCGCTGAGCGCTTGTCTCTGCGCTCA
Qy      788    CCGCGCGCGCGCGCGGACCGCGCGCGCGCGCGCGGAGTGTGAGCGCGCGGAACTCG
Db      29334  TCGCGGCTGTGAGGCGGTGCGCGCGCGCGCGCGCGCGCGCGTCTGCGCGCTACCTG
Qy      848    ACGGCGCTCTGCGGAGAACTCGCGCGGCTCATGGAGCTGCGCGCGCGCGGACGAGC
Db      29394  ACGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCG
Qy      908    ACTGCCAGAGCTGTTCGCGACCGCTTCACGCGACCGAC
Db      29454  CCGGCGCGGACGTCGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG

```

```

; SEQ ID NO 1
; LENGTH: 65140
; TYPE: DNA
; ORGANISM: Streptomyces noursei ATCC 11455
; US-10-203-295-1
Query Match      11.8%; Score 114.8; DB 19; Length 65140;
Best Local Similarity 46.9%; Pred. No. 1e-15;
Matches 449; Conservative 0; Mismatches 487; Indels 22; Gaps 2;

Qy      27    CCGTGGGCGGACGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
Db      19867  CGCGCGCGCGCTTACGCGGACCTGCGCGCGCATACACCGCGGCGCGCTGCTGCGGGA
Qy      87    CCGGACGCTGTGTGTGAGTCTGTACCGCGCGCGGAGACGCGCGGACGCTGCTGTGG
Db      19927  AGCGGTCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      147    CGAGTTCACCGCGCGCGGCGCGGACCTGCGCGCGCGCGCTTCCGAGACGCGGACTGCT
Db      19987  CGAGGACACCGTACCATGCGCGCTTACGACGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      207    GCTCGCGGTACAGAGAGAGTGGCGCTCAAGCGCGTGGCGCGCGCGCGCGCGCTGATCG
Db      20047  CTCCTGTGTCTCCGCGAGGTCCCGCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      267    GCGGCGCGGCTGCTGCGCGACCGCTGTGCTGCGCGAGCGGCGATGGCGCGGAGCTCGC
Db      20107  CGACTCTCTTTCAGGTGTGAGTGAACCGCGCTCAAGGCGCGCGCGCGCGCGCGCGCGG
Qy      327    GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20167  CACTGTGCGCTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      387    CATGACCGGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20227  CATCGGAGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      447    GCTGCGGCTGTGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20287  CGACTGTGTGTACACACCTTCAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      507    CCGGAGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20347  CGCGACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      567    CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20407  CGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      627    GGTGCTGTGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20465  GCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      667    CCGGAGTGTACGAGGACATCCAGCGGTCCACCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20525  CTGACCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      727    CTGCGGAGGCGCGTGTGCTCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20585  GCGAGCGACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      787    GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20645  GCGCGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      847    GAGCGGTGTGTGAGGAGTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Db      20705  ACCGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG
Qy      907    CACTGCGAGAGCTGTGTGCGACCGCTTCAACCGGACCGGACGAGGAGGAGGAGGCG
Db      20765  GTCCGCGACCTGTGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG

```

```
RESULT 7
US-10-156-761-7339
; Sequence 7339, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: SHIBATA, YOSHIOKI
; APPLICANT: SHIBATA, YOSHIOKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 7339
; LENGTH: 4569
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(4569)
US-10-156-761-7339
```

```
Query Match 11.3%; Score 109.2; DB 15; Length 4569;
Best Local Similarity 45.8%; Pred. No. 3.4e-14;
Matches 415; Conservative 0; Mismatches 488; Indels 3; Gaps 1;
```

```
QY 37 AGCGGGGCGGTGGCGGCGATGTTCCGCGGCTGCTCGCGAGGCGGCGAGCCGCGACGCTC 96
DB 1480 ACCGCGCCGACACCCGAGGCGATGCGCGCTCTGACGCGACGACCTCCGATCTC 1539
QY 97 GTCTGTGACTCTGTACCGCGCGCGGAGCGCGGAGCGCTGCTGTGTGGCGACGTCAC 156
DB 1540 GCGCGCGGTCTTCACGACGCGCGCGCGCGCGCGCGCGCGATCGGTGGCGCTCGCGCTC 1599
QY 157 GCGCGCGGCGCGGACGTCGCGCGCGCGCGCTCGCGGAGCGCGACGCTGCTGTGTGGCGCTA 216
DB 1600 GTGCGGCGCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1659
QY 217 CACGAGACGCTGCGCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 276
DB 1660 TGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1719
QY 277 CTGCTGCGCGACACCTGTCTCGCTCGCGAGCGCGGATGCGCGCGCGCGCGCGCGCGCGCG 333
DB 1720 AAGCTGCGCTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1779
QY 334 GCGCGCGGCGCTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 393
DB 1780 GCGCGCGACCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1839
QY 394 GCGCGACCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
DB 1840 GCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1899
QY 454 CTGCTGAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 513
DB 1900 CTGCGGCTGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1959
QY 514 AGCGGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 573
DB 1960 GCGCGCGGAAATCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2019
QY 574 CGCGTGGCGGTGACGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 633
```

```
DB 2020 GCGCTGAGCGTTCGCGGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2079
QY 634 CTGCGCGCTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 693
DB 2080 ACGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2139
QY 694 TCCAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 753
DB 2140 ATGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2199
QY 754 GCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 813
DB 2200 CTGATCTCGCAGATCTGCGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2259
QY 814 CATCCCGGCGGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 873
DB 2260 CTGCGCGCGATCAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2319
QY 874 CTCATGCGACCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 933
DB 2320 TGTCTCGCGACGACGCGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2379
QY 934 CACCGC 939
DB 2380 GTCCGC 2385
```

```
RESULT 8
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBATA, YOSHIOKI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
```

```
Query Match 11.3%; Score 109.2; DB 15; Length 9025608;
Best Local Similarity 45.8%; Pred. No. 3.3e-15;
Matches 415; Conservative 0; Mismatches 488; Indels 3; Gaps 1;
```

```
QY 37 AGCGGGGCGGTGGCGGCGATGTTCCGCGGCTGCTCGCGAGGCGGCGAGCCGCGACGCTC 96
DB 8781188 ACCGCGCCGACACCGGAGCGATGCGCGCTGTGACCGCGACGACCTCGCGATCTC 8781129
QY 97 GTCTGTGACTCTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 156
DB 8781128 GCGCGCGGTCTGACGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8781069
QY 157 GCGCGCGGCGCGGAACTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 216
```


Db 8781068 GTCGGGCCCAACCGGAAACGCTGGCCCTGGCCCCGCAAGCGCTGCGAAAGGCGCAAGG 8781009
QY 217 CACGAGGACGCTGCGCTCAAGGCGCGTGGCGCCCGTGAACCGCGGCTCATGCGCGCGCGCGG 276
Db 8781008 TGGCGGGGCGCGGCGCGACGCTGGCTTGTGCGGAGCCCGCTGCGGAGCCCGCGCGCGG 8780949
QY 277 CTGCTGCGCGCAACCTGTGCTGCTGCGGACGCGGCACTGCGCGG---AGCTGCGCGCGCAC 333
Db 8780948 AAGCTGCTTTCGCTTCCCGGCTGGAAGGCGGAGTTTCGAGCCAACTCGGAGGAGTC 8780889
QY 334 GCCCGCGCGCTGCGACGAGCTGGGCTTCAACCGATGTTGCGCCCGCGCGCGCGGCGATACC 393
Db 8780888 GCCCGGACCTTTCGCGCTGGCTGCTGCTGCGCGCGCTGCGCGCGCGCGCGCGCGAGCTG 8780829
QY 394 GCGCGACCGCTGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 453
Db 8780828 GCGCGGAGGAGGAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 8780769
QY 454 CTGCTGCGAGGCGCGCGCGCGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 513
Db 8780768 CTGCGGCTTCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 8780709
QY 514 ACGCGCGCGCAACCGAGCGCTGACGACGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 573
Db 8780708 GCGCGCGAAATCCAGCGACCGAGCGAGCTGACGCGCTTCTGCGCGCGCTTCTGACCGCG 8780649
QY 574 GCGCTGCGCGCTGACGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 633
Db 8780648 GCGCTGAGCTTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 8780589
QY 634 CTGCGCTTCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 693
Db 8780588 ACGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 8780529
QY 694 TCCAAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 753
Db 8780528 ATGATCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 8780469
QY 754 GCGCTGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 813
Db 8780468 CTGATCTGCGAGATCTGCGCGCTTCCGCTGCGCGCTTCCAGACCGCGATGCGAGCCCTAC 8780409
QY 814 CATCCCGGCGGATGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 873
Db 8780408 CTGCGCGCGGATCGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 8780349
QY 874 CTGATGCGAGCGCGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 933
Db 8780348 TGGTCGCGCGACGACCGTAGCCCGGTACCGGAGGAAACCGGCGGAGTCGCGACCTGTT 8780289
QY 934 CACCGC 939
Db 8780288 GTCGCG 8780283

RESULT 9

US-10-156-761-3440
; Sequence 3440, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIORITY FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30

; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIORITY FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3440
; LENGTH: 1884
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1884)
US-10-156-761-3440
Query Match 11.2%; Score 108.6; DB 15; Length 1884;
Best Local Similarity 45.5%; Pred. No. 5,9e-14;
Matches 436; Conservative 0; Mismatches 514; Indels 9; Gaps 1;
QY 6 CCGCTTCCCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 65
Db 855 CGCGCTCCCGGCGACCTTCAACGCGGACGCTGCAAGGCGGCTGACAGAGCGGCTGCGCTC 914
QY 66 GCTGCTGCGGAGCGCGGAGCGCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 125
Db 915 GGTCTTGGGACGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 974
QY 126 GCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 185
Db 975 CCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1034
QY 186 CCGGAGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 245
Db 1035 CGGCGTCCCGCGCTGCGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1094
QY 246 GCGCGTACCGCGCTATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 305
Db 1095 GGTGGGAGATGTCACCGCTGCGCGGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1154
QY 306 GGGCATAGCGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 365
Db 1155 CGCGCTCAACGCGGAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1214
QY 366 GATGTTGCGCGCGG-----CGCGCGGATGACCGCGCGCGCGCGCGCGCGCGCGCGG 416
Db 1215 GAGGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1274
QY 417 CACCAAGGACGCGCGCGCGCGCTACGCGCGCTGCTGCGCGCTGCTGCGCGCGCGCGCGG 476
Db 1275 CGCGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1334
QY 477 GCGCGTACCGCTACCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGG 536
Db 1335 GCGCTTCCGCGCGGCGGCTGCGCGCGAGGTTCCGCGCTGCGCGCGCGCGCGCGCGG 1394
QY 537 GCAAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 596
Db 1395 CGCGACCGGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1454
QY 597 CTGCGGCGGAGCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 656
Db 1455 CGGCTCGGTCACCGCGCGGAGGAGACCGGAGGTCGCGCTGCGCTGCGCGCGCGGAGAC 1514
QY 657 CGCGCGGACCGCGGAGGTAAGGAGACATCCAGCGCTTCAACCCCGCGCGCGCGCGG 716
Db 1515 CGTGTGCGCGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGTCG 1574
QY 717 GCGCGCGGCGCTGCGCGGAGCGCTGCGCTGCTGCGCGCGCGCTGCTGCGCGCGCGGAGAC 776
Db 1575 GAGACAGGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1634
QY 777 CCGTCCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 836
Db 1635 CCGCTGCTACGAGTGTGCGCGGAGCGTACGCGCTTCCGCGCGCGGCTGCGCGCGGCG 1694
QY 837 CGGGAACCTGACGCGGCTTCCGCGGAGACTCCGCGCGCGCTACGAGACCGGAGCTCCGCGG 896


```
Db      1695 CGACCCCTTCGCGCGGCGCTCTCGGCCAGGTGCGGCCCTTCCGCGAGGACCTGACCGG 1754
QY      897 GGGCAGCAGACACTGCGAGAGGCTGTTCCGACACCTTCCACCGACCGAGAGAGAGCGG 955
Db      1755 GACCTCGAAGTCCGCGGCGGCTCCGCTGCGCGGCGACGCGCTGACCGGCGTCCAGAGCG 1813

RESULT 10
US-10-472-587-5/c
; Sequence 5, Application US/10472587
; Publication No. US20040214274A1
; GENERAL INFORMATION:
; APPLICANT: YANAI, KOJI
; APPLICANT: SUMIDA, Naomi
; APPLICANT: WATANABE, Manabu
; APPLICANT: MORIYA, Tatsuki
; APPLICANT: MURAKAMI, Takeshi
; TITLE OF INVENTION: Transformants Producing Substance PF1022 Derivatives, Methods for
; FILE REFERENCE: 2003-1302A/MMC/00144
; CURRENT APPLICATION NUMBER: US/10/472,587
; PRIOR FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 82227/2001
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: Patent Ver. 2.1
; SEQ ID NO 5
; LENGTH: 969
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(966)
US-10-472-587-5

Query Match      11.1%; Score 107.4; DB 20; Length 969;
Best Local Similarity 47.7%; Pred. No. 1.3e-13;
Matches 378; Conservative 0; Mismatches 411; Indels 4; Gaps 2;

QY      21 CGTCGCTCTCGGCGGCGAGCGGCGGTGGCGGCAATTTGCGCGGCTGCTGCGGAGCG 80
Db      811 CGGGGCGGTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 752
QY      81 GGGCAGCAGCGACGCTGCTGCTGACCTGTAACCGCGCGCGAGCGCGGACCGCTGCGCT 140
Db      751 CGAAGAGCGCGAGGCGCTCGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 692
QY      141 GGTGGCGCAGTCAACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 200
Db      691 GGTGATGTCTCCGTAACCTCGGGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 632
QY      201 CGTCTGCTGCTGCGCTGACAGAGAGCGTGGCCCTCAAGCGCGCTGGCGCGCTG 260
Db      631 GCACTGTGTGG--GGCGCGGTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 574
QY      261 CATGCGCGCGCGCGCGCGCTGCTGCGCGCAACCTGTGCGTCCGAGCGGCGATGGCGCG 320
Db      573 GGGGAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 514
QY      321 GCTCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 380
Db      513 CGTCCGCTGCTGCTCTCCGCGCGTGAAGCGGTGAGGGGCTGCGCGCGCGCGCG 454
QY      381 CGCGCGCGATGACCGCGCGAGCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCG 440
Db      453 CGCGACAGAGGCGGTGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 394
QY      441 GGGCTGTGCGCGCTGCTGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 500
Db      393 GGTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 334
QY      501 GCAAGCAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 560
```

```
Db      333 GTGGGCGCGAGACTCGCGGCGCATGCGCTCCGAGCGAGCAGAGGTGTCCGCGAGCGCG 274
QY      561 GCTTCGCGCTTCGCGCGCGCTGCGGTGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCG 620
Db      273 GCGCGCGCGCATGAGCGCGGTGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCG 214
QY      621 --CAACAGATGCTGCTCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 678
Db      213 GGGAGACAGAGAGAGGTTCGCGGTCCCGAGAGGCGCGCGCGCGCGCGCGCGCG 154
QY      679 GGGAGATTCAGCGGTTCACACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 738
Db      153 GAGTTCGCCACACAGGCGAGCGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 94
QY      739 CTGCGCTCTTTCGCGCGCGCGCGTGTGCGCGCGAGCGCGCGAGCGCGCGCGCG 798
Db      93 CTGCGCGCTGCGCGCGCTTCCCGAGAGCGCGCGCGAGACATGCGCGCGCGCG 34
QY      799 GCGCAGCGCGCGCG 811
Db      33 GCGCAGCAGCAGC 21

RESULT 11
US-10-156-761-4723
; Sequence 4723, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; CURRENT FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4723
; LENGTH: 1707
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1707)
US-10-156-761-4723

Query Match      11.0%; Score 107; DB 15; Length 1707;
Best Local Similarity 48.0%; Pred. No. 1.3e-13;
Matches 342; Conservative 0; Mismatches 365; Indels 6; Gaps 1;

QY      238 GCGTGGGCGCGCGTGAACCGGCTCATGCGCGCGCGCGCGCGCGCGCGCGCG 297
Db      820 GCGGCGCGCGGTGAGAGGCGGTACGCGGTGCGCGGTGCGCGGTGCGAGAGAG 879
QY      298 GTCCGAGCGGAGATGGCGCGCGAGCGTCCGCGCGCGCGCGCGCGCGCGCG 357
Db      880 TCCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 939
QY      358 CTCACCCGATGTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 417
Db      940 GGGGTTCAGAGGCGCACCGGCTTTCCTGATGAGAGAGCGCGCGTCCCGGATGAC 999
QY      418 ACCAGGAGCGGCGCGCGCGCTCAACGCGCGCTGCTGCGCGCTGTGAGAGCGCGCG 477
Db      1000 GCGGCGACCTTGAAACAGCAGCGCGCGCTGAGTCCGAGCGGGAACGTGACCGCGCGAC 1059
```

[illegible]

```

RESULT 12
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      11.0%; Score 107; DB 15; Length 9025608;
Best Local Similarity 48.0%; Pred.No. 9.5e-15;
Matches 342; Conservative 0; Mismatches 365; Indels 6; Gaps 1;

QY      228  GCCCTGCGCCCGCTACCCGCGCTATCGCGCGCGCGCGCGCGCTGCTGCGCCGACACCTCTGTC 297
DB      5778628  GCGCGGCGGCTGACGGGCGTACCGGTAACGCGCGCTGCGCTGCGCTGACGAAACCCCGCC 5778668

```

Oy	298	TTCCGAGACGGGCAATGGCCGGGAGAGTGGCGGdCAGCCAGCCCGGCGTCCAGAAAGTGGG	357
Db	5778688	TTCCGGGCGGGGCGGGCCTTCGCCCCCGCGGTATCCAGCAGAGGGCTGCCCCCTTCGACTCTCC	5778747
Oy	358	CTCAACCCGATGTGTCCGCCCCCGCGCGCCGGATGACCGGCCGACCCGTGGCGCGTGTCTC	417
Db	5778748	GCGGTTCAGGGGCGCCAGCGGGCTTCTCCGTATGGAGAGGCGCGGATCGCCGGAGTGAC	5778807
Oy	418	ACCAAGGAGACGGGAGCGGGCGTTCACGGGCCCCCTGTGCGGTCTGTCCAGGGCGCGCGGGGACG	477
Db	5778808	GCGGCGCAACCTTGGAAACAGCAGCGCGGCTTGGATGCAACGGGGAACGTTCAGCCGCGGAC	5778867
Oy	478	CCCGTACCGGCTCAGCGCGGAGAGACGACCGGACGAGCGGCGCACCCAGGCGCTTAGC	537
Db	5778868	AGCGCGCGCAAGCCCGGAGTTCAGGCGCGAGCTATCATCCGCGCTCGCGCCGACGGTGGC	5778927
Oy	538	CACGCGGTGTCTCTTCTTCGAGGTTCGCCCTCGCCGCTCGGCGTTCGACGTCCGAGC	597
Db	5778928	GCCGCGGCGCGGTCTCCCGAGCGCGCGCGCGCCGAGCGCCGTGTCGAGGCGCCCCCGGCC	5778987
Oy	598	CTGGCGGCGAGACCGACCGCGCGCCCGCAACAGATGTGCTGCGCCCTCAGGGCCGATGTCTC	657
Db	5778988	GCGCGCGCGCTCGACTGTCTGACCTTCGATGCGATGCGCCGTGACATTCAGCGCGCGCTCAC	5779047
Oy	658	GGCGGCAAGCCCGGAGGTATGACGGGAGCATTCAGCGATTCAAACCCCGCGGCGCGTTCGCG	717
Db	5779048	CGGCTGTTCGGGGGTGGGGCTGTGTGACATCGAATCG-----CTCGCGAGGGCTTCGCG	5779107
Oy	718	CGCGCGGCGCTTCGCGGAGCGCTTCGCGCTCTTTCGCGCGCGCTGTGTGGCGAGACCGGAC	777
Db	5779102	GACCGCGCGATGGCGCGCGCGACGTGAGCTTGTGGATGGCGCGATGCTTCGCGCGAGAGTGGC	5779161
Oy	778	CGTCCGACGAGCCCCCGGGCGCGCGACGCGCCCGCGCATCCCGGGGATGCGACGCGCGC	837
Db	5779162	GCTTCGCGCGCGCGCGCGCGCGCGCGCGGACATCACCCCGCTGTCTGCGCTCGCGACG	5779221
Oy	838	GGGAACCTCGACGGCGTCTTCGAGGAATCCGCTGAGCTCATGGAGCCGAGACTTCGCGAGC	897
Db	5779222	ATGGCGCCCGACGTCTGTGGAACGATATCGCGGGCTTCGACGAGCGGCTGCGCGGCTTC	5779281
Oy	898	GGCCAGGACCATCTGCCAGAGCTGTTCCGACCTTCACCGGACCGGACGACGACGA	950
Db	5779282	GACGAGGAAGCAGCGCGCGAGATATCCAGACCGTGGCGCGCGCTGTCTGACAA	5779341

```

, RESULT 13
, US-10-126-927-68/C
, Sequence 68, Application US/10126927
, Publication No. US20030082575A1
, GENERAL INFORMATION:
, APPLICANT: The Scripps Research Institute
, APPLICANT: Schultz, Peter G
, APPLICANT: Wang, Lei
, APPLICANT: Anderson, John C
, APPLICANT: Chin, Jason
, APPLICANT: Liu, David R
, APPLICANT: Megliery, Thomas
, APPLICANT: Meggers, Eric L
, APPLICANT: Mehl, Ryan A
, APPLICANT: Pastnak, Miro
, APPLICANT: Santoro, Stephen W
, APPLICANT: Zhang, Zhiwen
, TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
, FILE REFERENCE: 54-000120US
, CURRENT APPLICATION NUMBER: US/10/126,927
, CURRENT FILING DATE: 2002-04-19
, PRIOR APPLICATION NUMBER: US 60/285,030
, PRIOR FILING DATE: 2001-04-19
, PRIOR APPLICATION NUMBER: US 60/355,514
, PRIOR FILING DATE: 2002-02-06
, NUMBER OF SEQ. ID NOS: 79
, SOFTWARE: PatentIn version 3.1
, SEQ ID NO 68

```

LENGTH: 3305
TYPE: DNA
ORGANISM: Streptomyces venezuelae
US-10-126-927-68

Query Match 11.0%; Score 106.2; DB 14; Length 3305;
Best Local Similarity 48.2%; Pred. No. 1,6e-13;
Matches 391; Conservative 0; Mismatches 413; Indels 7; Gaps 3;

```
21 CGTCGCTGCTGCGGCGAGGCGGCTGGGCGGCGCATGTTGCGGGGCTGCTGCGGAGCG 80
2817 CCGGGCTGCTGCGGCGGCTGCGGCTGCTGCGGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2758
81 GGGCAGACCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140
2757 CGAAGAGGCGGAGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2698
141 GGTGGCGGAGCTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 200
2697 GCTGGAGTGTCCCGTACACCTCGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2638
201 CGTCGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 260
2637 GCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2580
261 CATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 320
2579 GCGGAGGCGGAGGCGGCGGAGGAGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
321 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 380
2519 CGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2460
381 CGCGGCGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 440
2459 CCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2400
441 GCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
2399 GGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340
501 GCAACGACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 560
2339 GTGGGCGGCGGAGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2280
561 GCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
2279 GCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220
621 --CAACAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678
2219 GCGGAGCAGAGAGAGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2160
679 GGGGACATGACGAGGCTGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 738
2159 GACGTGCGCAGCAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2103
739 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 798
2102 GAGCGTGGCGGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2043
799 GCCGACGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 829
2042 GCCGCGGAGCAGCAGCGCTCGGGGGAGAGCGG 2012
```

RESULT 14
US-10-126-927-67/c

Sequence 67, Application US/10126927
Publication No. US20030082575A1
GENERAL INFORMATION:
APPLICANT: The Scripps Research Institute
APPLICANT: Schultz, Peter G

```
APPLICANT: Wang, Lei
APPLICANT: Anderson, John C
APPLICANT: Chin, Jason
APPLICANT: Liu, David R
APPLICANT: Magliery, Thomas
APPLICANT: Meggers, Eric L
APPLICANT: Mehl, Ryan A
APPLICANT: Pasternak, Miro
APPLICANT: Santoro, Stephen W
APPLICANT: Zhang, Zhien W
TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids
FILE REFERENCE: 54-000120US
CURRENT APPLICATION NUMBER: US/10/126,927
PRIOR FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,030
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: US 60/355,514
NUMBER OF SEQ ID NOS: 79
SOFTWARE: Patentin version 3.1
SEQ ID NO 67
LENGTH: 12391
TYPE: DNA
ORGANISM: Plasmid pSC101, Streptomyces venezuelae papABC
US-10-126-927-67
```

Query Match 10.9%; Score 105.4; DB 14; Length 12391;
Best Local Similarity 48.2%; Pred. No. 1,5e-13;
Matches 389; Conservative 0; Mismatches 411; Indels 7; Gaps 3;

```
21 CGTCGCTGCTGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 80
3171 CCGGGCTGCTGCGGCGGCTGCGGCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 3112
81 GGGCAGCGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 140
3111 CGAAGAGCGGAGGCGGCTTGGGAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTT 3052
141 GGTGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 200
3051 GCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2992
201 CGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 260
2991 GCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2934
261 CATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 320
2933 GCGGAGGCGGAGCGGCGGAGGAGATCAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2874
321 GCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 380
2873 CGTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2814
381 CGCGGCGATGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 440
2813 CCGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2754
441 GCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500
2753 GGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2694
501 GCAACGACGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 560
2693 GTGGGCGGCGGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2634
561 GCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 620
2633 GCGCGGCGGCGGAGCGGCTGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2574
621 --CAACAGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 678
2573 GCGGAGCAGAGAGGCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2514
```

Sequence	Position
679	GGGGGCAATCGACGGTTCACACCCCGGGCGGGTCTCGGCGCCCGGGGCTGCGGAGGCG 738
Db	2513 GACGTGCGCCACCGAGCAGGCGGTCTCGGCG --CGTCCGGCGCGCGGTACGAGGTGCACAC 2457
Oy	739 CTGGCGCTCTTCGCGCGCGCTGCTGCGCGACGACCCGCGACCGTGCACGACCCCGGGCGC 798
Db	2456 GAGCGCTGCGGCTGCGCCCGCTCCCGCAGCAGAGCCCGGCGAACATGCGCCACCGCTCCGCT 2397
Oy	799 GCCGACGCGCCCGGCGCATCCGGGGGA 825
Db	2396 GCCGCGCGACGACGACCTCCGGGGGAA 2370

```

RESULT 15
US-10-156-761-6049
: Sequence 6049, Application US/10156761
: Publication NO. US20030119018A1
: GENERAL INFORMATION:
: APPLICANT: OMURA, SATOSHI
: APPLICANT: IKEDA, HARUO
: APPLICANT: ISHIKAWA, JUN
: APPLICANT: HORIKAWA, HIROSHI
: APPLICANT: SHIBA, TADAYOSHI
: APPLICANT: SAKAKI, YOSHIYUKI
: APPLICANT: HATTORI, MASAHIRA
: TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
: FILE REFERENCE: 249-262
: CURRENT APPLICATION NUMBER: US/10/156, 761
: CURRENT FILING DATE: 2002-05-29
: PRIOR APPLICATION NUMBER: JP 2001-204089
: PRIOR FILING DATE: 2001-05-30
: PRIOR APPLICATION NUMBER: JP 2001-272697
: PRIOR FILING DATE: 2001-08-02
: NUMBER OF SEQ ID NOS: 15109
: SEQ ID NO 6049
:
: LENGTH: 1167
: TYPE: DNA
: ORGANISM: Streptomyces avermitilis
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)..(1167)
:
: US-10-156-761-6049

```

Query Match	10.9%	Score 105.2	DB 15	Length 1167
Best Local Similarity	46.1%	Pred. No. 3.5e-13		
Matches 393	Conservative	0	Mismatches 453	Indels 6
				Gaps 1

OY	2	TGAGGCGCTTCCCCCGGCGAGCGTCGTCGTCGCGGAGCGGGGCGGGTGGGCGGCAATGTTGCG	61
Db	104	TCAGGCGCGGACCGGACCGGACACCTCGGGTGCCTCGGCGGTACCGAGCGGGCGGGGCC	163
OY	62	CCGGGCTGCTGCAGGAGGGCGGGGACGCGACGCTGTCGTGACCTCGTACCGCGCCGG	121
Db	164	CCGTGAGAGCGGACCGGGGCGCTGACCGGCTGGCGCGCCCGGAGCGCGCGGACCGCTTCCGCG	223
OY	122	GACGGCCGGAGCGCTGCTGTGGGCGAGCTACCGCGGCGGGGCGGAACTTCGGGCGCG	181
Db	224	CGGACCTTCACGCGCTTCCAGGACCGGCGCGCGACCGCGCTCGACTTCTCGGGCGCGGCC	283
OY	182	CCCTTCGGGAGCGGGAGCTCGTCTGTGCTGCGCGCTGACAGAGAGAGTGGCCCTCAAGGCGCG	241
Db	284	GCTTGAGAGCGCGAGGAATTCACCCCCCTCCCGGGCGGAGGAACTCGCCCTGTCGCGCG	343
OY	242	TGGGCGCCGTGACCCCGGCTCATGCGGGCGGGGCGGCGCTGTGGCGGACCCGTGTCGCTCC	301
Db	344	CCCTGCGGAGCGCGCGCTCTTCCGAGGCGGTACGACTCTCTGTGTGCACTTCCGCCGG	403
OY	302	GGAGCGGAGATGGCCCGCGAGCTCGCGGAGCCACGCGCCCGGCGGTCCAGGACGCTGGGCTTCA	361
Db	404	CCCCGCAAGCGCTGCGCCTCTCTGCGCTCTCCCGAGGAGACTGCGCGCTATCTGCGCGCGC	463
OY	362	ACCGGATGTCGCCCCCGCGCGCGCGGACATGACCGGCGGACCCGTGGCGCGCTGTGTACCA	421

Db 464 TCTTCCCCGCCGAAGCCAGAGCGAGCCCGCGCTCTGCGCCCTGCTGCGGCTCGCCG 5233

Qy 422 GGGACGAGCGCGAGCGTCAAGCGCCCTGCTGCGGCTCTGTCAGAGGCGAGCGCGAGCGCCG 4811

Db 524 GCGTCCCGATGCGCCGCGGACTGGCTGTACGAGAGCGGCCGCCCGCTGGGAGCTGGAGCTCG 5833

Qy 482 TACGCGTCAAGCGCGAGAGACACGACCGGAGAGAGCGGCGCCACCCAGGCCCTGACGTCAG 54111

Db 584 CCGCGGTTCAGAGCGGCTGTGCGAGAGCCCGGCAAGACCTGTACGACTGTGTCGCGAGGCCG 6433

Qy 542 CCGTGTCTCTCTTTCGAGGCTCGGCGCTCGGCGCGCTCGAGGCTGAGGTCGCGAGCCCTGCG 6011

Db 644 GACCGGCGCGGCCCGACGCGCGCTCGCGCGCGCGCGCGCACCGAGCTTGTGCGCTCGCGCGCTGC 7033

Qy 602 CGGCGAGCGGACCGCGCGCCCAACAGAGTGTCTTCGCGCTTCGAGCCGCTGTGCTCGAGC 6611

Db 704 GCATCGACACCTGTATCGGAAACGGGCTCTGCGCCGAGAGAGCGACGACGCGCTGTGCGG 7633

Qy 662 GCAGCGCCCGAGGTGACGAGGACATTCAGCGGCTCCAAACCCCGGCGCGGCTTCGCGCGCC 7211

Db 764 CCGGCGCTTCGCGCGCCAGACGCGCAAGTTCCTTCGAGAGGTGCGGAGGAGTACGAGCTTC 8233

Qy 722 GGGCGCTCGCGAGGCGCTGCGCTCTTCGCGCGGCGTGTTCGAGGAGACGACCGGACCGTG 7811

Db 824 ACGAATCT-----CGGACCTTCGAGACGACCGCGCGGCGGAGACGACTGGCGGCAC 8777

Qy 782 CCGAGCGCCCGGAGCGCGCGACGCGCCCGGCAATCCCGGAGATGCGAGCGCGCGCGAGA 8411

Db 878 TCGCGGTACCCCGGCGTCAACGACGCGCTGTGGCGGAGTGAAGTGTCTGTCGCGCGACCGCC 9377

Qy 842 ACCTCGACGCGC 853

Db 938 TCGCTGACGACG 949

Search completed: October 5, 2005, 09:57:35
Job time : 982.313 secs

22

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: October 4, 2005, 20:20:08 ; Search time 3664.34 Seconds
(without alignments)
10065.737 Million cell updates/sec

Title: US-10-089-514-5

Perfect score: 969
1 atgcgcgcgtcccccgcag.....aagcgagagaagcagcatga 969

Sequence: IDENTITY_NUC

Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

EST: *
1: gb_est1: *
2: gb_est2: *
3: gb_hic: *
4: gb_est3: *
5: gb_est4: *
6: gb_est5: *
7: gb_est6: *
8: gb_gest1: *
9: gb_gest2: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	134.6	13.9	1462	AG441877	Mus muscu
C 2	133	13.7	1469	AG365356	Mus muscu
C 3	131.2	13.5	1610	B2569386	pac82-164
C 4	128.6	13.3	1406	CG756569	P051-4-B0
C 5	127	13.1	2243	AG381986	Mus muscu
C 6	126.8	13.1	1137	BG809979	mgct002xd
C 7	126.8	13.1	1350	BM450237	AGENCOURT
C 8	126.6	13.1	1780	AG448243	Mus muscu
C 9	126	13.0	1198	CC190008	CH261-138
C 10	125.8	13.0	1692	AG396765	Mus muscu
C 11	125.2	12.9	1073	CC209412	FGAS02117
C 12	125.2	12.9	1821	CL090560	ISB1-17N1
C 13	124.6	12.9	1218	CL081011	CH216-160
C 14	124.6	12.9	1267	CL476590	SATL_258
C 15	124.4	12.8	2332	AG363333	Mus muscu
C 16	123.8	12.8	1355	AG429718	Mus muscu
C 17	123.6	12.8	1569	AG341503	Mus muscu
C 18	123.6	12.8	2332	AG363333	Mus muscu
C 19	123.4	12.7	1878	AG393083	Mus muscu
C 20	123.2	12.7	1452	AG032979	Pan trogl
C 21	123	12.7	982	BQ687717	AGENCOURT
C 22	123	12.7	1281	BG852363	1024034A0
C 23	122.8	12.7	1070	B2557810	pacs1-60
C 24	122.8	12.7	1359	CL507830	SATL_788

25	122.8	12.7	1421	8	B2569488	pac82-164
26	122.4	12.6	1280	9	CL470317	SATL_140
C 27	122	12.6	2017	9	AG435283	Mus muscu
C 28	121.4	12.5	1280	9	CL470317	SATL_140
C 29	121.2	12.5	1956	9	CG754548	P050-1-A1
C 30	120.8	12.5	1885	9	AG363563	Mus muscu
C 31	120.8	12.5	1628	9	CG757066	P052-2-A0
C 32	120.8	12.5	1822	9	AG435170	Mus muscu
C 33	120.6	12.4	1216	9	AG441574	Mus muscu
C 34	120.6	12.4	1300	9	AG430756	Mus muscu
C 35	120.4	12.4	1131	9	AG042920	Pan trogl
C 36	120.4	12.4	1552	9	AG430101	Mus muscu
C 37	120.2	12.4	1339	4	BM458211	AGENCOURT
C 38	120.2	12.4	1853	9	AG382854	Mus muscu
C 39	119.6	12.3	1625	9	AG043477	Pan trogl
C 40	119.4	12.3	1116	8	B2569478	pac82-164
C 41	119.4	12.3	1360	9	AG136785	Pan trogl
C 42	119	12.3	1360	9	CL469817	SATL_133
C 43	118.8	12.3	885	8	AG159162	Pan trogl
C 44	118.8	12.3	965	8	AO893417	HS 4832 A
C 45	118.6	12.2	1674	9	CL078342	CH216-149

ALIGNMENTS

RESULT 1
AG441877/c
LOCUS
DEFINITION
Mus musculus molossinus DNA, clone:MSMG01-323H09.TJ, genomic survey
sequence.
ACCESSION
AG441877
VERSION
AG441877.1
KEYWORDS
GSS.
SOURCE
Mus musculus molossinus
ORGANISM
Mus musculus molossinus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSMG01
TITLE
Unpublished
JOURNAL
2 (bases 1 to 1462)
AUTHORS
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
DIRECT SUBMISSION
SUBMITTED (17-NOV-2003) Masahira Hattori, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suenho-chou, Tsukuba, Iku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/
Tel:81-45-503-9111, Fax:81-45-503-9170)
Clones are derived from the mouse BAC library MSMG01. For BAC
library availability, please contact Kunya Abe (abe@rtc.riken.jp).
Tsukuba Institute, Bio Resource Center,
The Institute of Physical and Chemical Research (RIKEN) 3-1-1
Koyadai, Tsukuba, 305-0074 Japan
phone: 81-298-36-9189, fax: 81-298-36-9199
e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector : pBAC3.6
R Site 1 : EcoRI
R Site 2 : EcoRI.
Location/Qualifiers
1. 1462
/organism="Mus musculus molossinus"
/mol_type="genomic DNA"
/sub_species="molossinus"
/db_xref="taxon:57486"
/clone="MSMG01-323H09.TJ"
/sex="male"
/tissue type="mixture of kidney and spleen"
/clone_id="MSMG01 Mouse Male BAC library"

ORIGIN

Query Match	13.9%	Score 134.6;	DB 9;	Length 1462;
Best Local Similarity	50.7%;	Pred. No. 6e-13;		
Matches 422;	Conservative 0;	Mismatches 396;	Indels 14;	Gaps 4;

[illegible]

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

Bakayocsa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Scutogomathi; Muridae; Murinae; Mus.
1
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
BAC end Sequences of Library MSWG01
Unpublished
2 (bases 1 to 1459)
Hattori, M., Toyoda, A., Noguchi, H., Kojima, T. and Sakaki, Y.
Direct Submission
Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical

FEATURES

Source

ORIGIN

Query Match 13.7%; Score 133; DB 9; Length 1469;
Best Local Similarity 47.8%; Pred. No. 1.1e-12;
Matches 408; Conservative 0; Mismatches 442; Indels 4; Gaps 1;

OY	9	CTTCCCCCGAGCGCTCGTCTGTGCGCGAGCGGGGCGGTGGCGGCATGTTTCGCGAGGCT	68
Db	1384	CTCGCTCTGCTGCGCGCTCTCGCGGCGCGGCGCCCGCTCTCGGCGCGCGCGCGCGCG	1322
OY	69	GCTGCGGGAGCGCGGCGAGCGCGACGCTCGTGTTCGACTTCGACCGCGCGGAGCGAGC	128
Db	1324	CNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGCGGCGGAGCGCGCGCTCCCGGCGCGC	126
OY	129	GGAGCGCTGCGCTGAGTGGGCGACGTGACCGAGCGCGGAGCGCGACGTGCGGCGCGCTTCG	188
Db	1264	CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	120
OY	189	GGAAGCGGACCTCTGCTCTGCTTCGCGGTACAGAGGACGTGGCCCTCAAGGCGGTGGCGCC	248
Db	1204	CGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1145
OY	249	CGTGAACCCGAGCTCATGCGGACCGGCGCGCGCTGTGCGCGACACCTGTTCGTTCGAGCGG	308
Db	1144	CCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1085
OY	309	CATGAGCGCGGAGCTGCGCGGCGCAAGCGCGCGCGCGCTCCAGACGTTGGGCTCAACCCGAT	368
Db	1088	CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1022
OY	369	GTTGCGCGCGCGCGCGCGCATGACCGGCGGACCGCTGGCGCGCGTGTGTCACAGAGGACG	428
Db	1028	GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	969
OY	429	GCCGAGCGTCAAGCGCGCTGCTGCGGCTTCGTGAGAGGCGCGGCGGCAAGCGCTGTACGCGCT	488

	SOURCE	ORGANISM
	Pristionchus pacificus	Pristionchus pacificus
	Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;	
	Neodiplogasteridae; Pristionchus.	
REFERENCE	1 (bases 1 to 1406)	
AUTHORS	Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.	
TITLE	An integrated physical and genetic map of the nematode Pristionchus pacificus	
JOURNAL	Mol. Genet. Genomics 269 (5), 715-722 (2003)	
MEDLINE	22835951	
PUBMED	12884007	
COMMENT	Contact: Sommer RJ Evolutionary Biology Max-Planck-Institute for Developmental Biology Spemannstr. 37-39, Tuebingen D-72076, Germany Tel.: 00497071601371 Fax: 00497071601498 Email: ralf.sommer@tuebingen.mpg.de Class: BAC ends.	
FEATURES	Location/Qualifiers	
source	1..1406	
	/organism="Pristionchus pacificus"	
	/mol_type="genomic DNA"	
	/strain="California"	
	/db_xref="taxon:54126"	
	/clone_lib="Epa SCRI BAC Library"	
	/note="The library was generated by a partial digest of the genomic DNA with EcoRI and cloning into the BAC vector."	
ORIGIN		
Query Match	13.3%; Score 128.6; DB 9; Length 1406;	
Best Local Similarity	47.4%; Pred. No. 6e-12;	
Matches	427; Conservative 0; Mismatches 471; Indels 3; Gaps 2,	
OY	49 GCGGCGAATTTGGCCGGGCTGTGGGGAGAGCGGSCAACGCCACTGTCTGCACCTC 108	
Db	421 GGCGGCCCCGNGNCCCCCGCGCGGCGGCGCCCGCCCGCCCGCGCGGNCGCC 480	
OY	109 GTACCAGCCGCGGAGACGGCCGGACCGCACTGACCTGTGTGGAGCGAGTACACGCGCGGAGGCC 168	
Db	481 GCGGCGCCCGCGGNCCTGGCGGGAGCGCGGCCCGCCCGCGGAGCGGCGCGCGGCCCG 540	
OY	169 GAATTCGGGCGCCCTTCGGGAGCGCGAATTGTCTTGCTTGCCGTACAGAGACGTG 228	
Db	541 CGGGGGCCCCCGCGCGCGCGCGGAGGAGCGCCCGCGCCCGCCCGCCCGCCCGCCC 600	
OY	229 GCCCTCAAGGCGCTGGCCGCCCTGTACCCGCGTCAATGCGGAGCGGCGGCTGTGGCCGAC 288	
Db	601 GCCCGCGCGCGCGCGCGCCCGCCCGCGCGCCCGCGCGCCCGCCCGCGCGCGCGCC 660	
OY	289 ACCCTGATCCGATCCGAGACGAGCATGAGCGCGAGACTGCGGCGCCACGCGCCCG-GGATCA 347	
Db	681 CCGCCCCCCCCCGCCCCCGCGGCGCCCGCGGCGCCCGCGGCGCCCGCGCGCGCGCC 720	
OY	348 GCACGTGAGGCTCAACCCGATGTTGCCCGCGCGCGAGCAATGACCGAGCGAACCCGTGAC 407	
Db	721 CGCGCGCCCCCCCCCGGGGAGCGCGCCCGCCCGCGCGCCCGCGCGCGCGCGCCCG 780	
OY	408 CGCGGTGATCACAGGAGACGGGCGCGGCGGTACAGGCGCTGTGCGGCTGTGTGAAGGGCG 467	
Db	781 CGCGGCGCGCGCGCGGAGGAGCGCCCGGCGCCCGCGCGCCCGCGCGCGCGCGCGCC 840	
OY	468 CGGCGGACGAGCGCGTACAGGCGGAGAGAGACACCGGAGAGAGGAGGCGGACCA 527	
Db	841 CGCGCGCCCCCCCCCGCGCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900	
OY	528 GAGCCTGACGACAGCGCGTGTCTTCTCTTGAGAGCTGCGCCCTGCGCCACTGAGGTGCA 587	
Db	901 CCGCGGCGCCCGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 960	
OY	588 GGTCCGGGCGCTGGCGGCGAGCGACCGGCGCGCCCAACAAGGTGCTTGCCCTTCGAGC 647	

[illegible]

QY 419 CCAGGAGAGGAGCGGGGCGTCAACGAGCGCTGACGAGCGCGCGCGCGAGC 478
 Db 662 GGGGAGCGGCGCGCGCGAGCGCGGGGCGGGGCGGCCCGCGCGAGGAGGCGGCGCG 603
 QY 479 CCGTACGAGCTCAACGCGCGAGAGGACGACCGGACCGACGAGCGCGACCGAGCGCTCGAGCG 538
 Db 602 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 543
 QY 539 ACGCGGTGTCTCTCTCTTTCGGGCTTCGCCCTTCGCGCGCTTCGGCGTTCGACTCGGAGCGCC 598
 Db 542 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 483
 QY 599 TGGCGGAGCGAGCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 658
 Db 422 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 423
 QY 659 GCGGAGCGCGCGAGGTGTACGAGGACATTCAGCGAGTTCGACCGCGCGCGCGCGCGCGCG 718
 Db 422 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 363
 QY 719 GCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 778
 Db 362 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 303
 QY 779 GTTACCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 838
 Db 302 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 243
 QY 839 GGAACCTTCGACGCGGCTCTTTCGGGAACTTCGCGCGGCTCATTTGGAACCGAGGCTCGC 893
 Db 242 CNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 188

RESULT 7	BM450237	1350 bp	mRNA	linear	EST 05-FEB-2002
LOCUS	AGENCOURT_6393396 NIH_MGC_72	Homo sapiens	CDNA clone	IMAGE:5528315	
DEFINITION	5', mRNA sequence.				
ACCESSION	BM450237				
VERSION	BM450237.1	GI:18499277			
KEYWORDS	EST.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS	Mammalia; Euteleostomi; Primates; Cetartiodactyla; Homnidae; Homo.				
TITLE	1 (bases 1 to 1350)				
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .				
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)				
	Unpublished (1999)				
	Contact: Robert Strausberg, Ph.D.				
	Email: cgapbs-r@mail.nih.gov				
	Tissue Procurement: ATCC/CDTDF/DPF				
	CDNA Library Preparation: Life Technologies, Inc.				
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)				
	DNA Sequencing by: Agencourt Bioscience Corporation				
	Clone distribution: MGC clone distribution information can be				
	found through the I.M.A.G.E. Consortium/LNLN at:				
	http://image.llnl.gov				
	Plate: LLM12204	row: p	column: 12		
	High quality sequence stop: 370.				
FEATURES	Location/Qualifiers				
source	1..1350				

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="IMAGE:5528315"
/tissue_type="melanotic melanoma"
/lab_host="DH10B (phage-resistant)"
/clone_idb="N1H MGC 72"
/notes="Organ: skin; Vector: pCMV-SPORE6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: oligo dt.
Average insert size 2 kb. Library constructed by life

```

Technologies."

Query Match	13.1%;	Score 126.8;	DB 4;	Length 1350;
Best Local Similarity	43.8%;	Pred. No. 1.2e-11;		
Matches 399;	Conservative	0;	Mismatches 507;	Indels 5;
				Gaps 2

QY	42	GGCGGTGGAGGAGCAATGTTTGGCGGGGTGTGCGGGAGGCGGACGCGACGCTGCTGT	101
Db	286	GCTGTGTTCGGATCCATCTCTCTTTCTTCAGAGCTCTCCCTCGAGAGTGGATGTGTGG	345
QY	102	CGACTCTGTACCGCGCGCCGGAGACGGCCGGAAGCTGTGTGGGCGAGTCAACGCGCC	161
Db	346	TGGGACCGGAGCGGGGCGGANNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN	405
QY	162	GGGGGTCGAATCTCGCGGCGGCTTCGCGGGAAGCGGACCTTGTCTCTGTCTGCGCTACAGA	221
Db	406	GCGCCCCCGGAGCCCGGGGGCGCCCGCGCCCTCGGGGCTCGCGCCCGCCCGCCCGG	465
QY	222	GGAAGCTGAGCCCTCAAGGGCGGTGTGGCGCCGCTGACCCGAGCTCAATGCGAGCGGGGCGCTGCT	281
Db	466	GNGCGCCCCCGGGGGCGCCCGGGGGCGCGGGCGCGGGGGGACACNNNNNNNGNNGGCGNG	525
QY	282	CGCGGACACCTCTGTCTGTCTCGGAGCGGAGCATGGCGCGGAGCTCGCGGCCACGCCCCGG	341
Db	526	GNGCGNCNMGAGNAGGAGGGAGNNNGAGGGGGGGGGCGCCCGGGGCGGGGCGCGCCCGCCCGNG	585
QY	342	CGTCAGACAGTGGGGCTTCAACCCGAGATTGTGACCCCGCGCGGGCATGACCGGCGGAC	401
Db	586	CCCCCCCCCGGCGCCCCCGGGCCCGGAGCCCGAGNCGAGCGGTGGAGCCGCGGCGCNCGCGC	645
QY	402	CGTGGCCGCGCTGTGTCAACAGAGGAGCGGGCGGCGTCAAGGCCCTGTCTGTGTGTGCA	461
Db	646	GGAGNCCGCGCCCCGGGGCGGCGCCGCGCCCGGCGCGCGCGGAGNGGCGCCGGAGCGC-C	704
QY	462	GGGGGCGCGCGGAGCGCCCGGTACCGGCTCAAGCGGAGAGACGACACCGGACGAGCGGCG	521
Db	705	GGGCCCCCGCGGGGCGCCGCGAGNAGNCGCGCGCGCGCCGCGCGCGCGGGGGGGCGCC	764
QY	522	CACCCAGGCGCTGTGACGACAGCGCGGTGTCTCTCTTCTGTGGAGCTGCGCTGTGCGCGCTGTG	581
Db	765	CGGCGCGCGCCGAGCCCGCGCGGAGNCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	824
QY	582	CGTGCAGCTCCG---GAGCCTGGGGGCGAGGACCGCGCGCCACCAAGGTGTGTCTG	637
Db	825	CGCGCGCCCCCGCGCGGCGCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	884
QY	638	CCCTTCGTGACCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	697
Db	885	CCCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	944
QY	698	ACCCCGGGGCGGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	757
Db	945	GAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1004
QY	758	TGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	817
Db	1005	CCCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1066
QY	818	CGGGGAGATGCGACGCGCGCGCGGAACTTCGACGCGGTCTTCTGGGAGAACTTCGCGCGGCTCA	877
Db	1065	CCCCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1122
QY	878	TGGGACCGGAGACTCGCGCGCGGCGGACGACGACCTGCGAGAGAGCTGTTCGCGACCTTCAC	937
Db	1125	CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1184
QY	938	GCACCGACGAC 948	
Db	1185	NGGCGNCCGCC 1195	

[illegible]

JOURNAL
 COMMENT

Unpublished (2003)
 Contact: Mm L Crosby
 Bioinformatics
 University of Saskatchewan, Department of Computer Science
 1C101 Engineering Building, 57 Campus Drive, Saskatoon,
 Saskatchewan, S7N 5A9, Canada
 Tel: 306 966 1769
 Fax: 306 966 2033
 Email: fgas_estes@cs.usask.ca
 This sequence is the direct result of the Base calling software
 Phred (default parameters). It is the raw base calls. To aid in the
 identification of the high quality insert the software Lucy
 (default parameters) has been run on this sequence. Lucy identified
 the region 120,1951.
 Plate: L5B017 row: G column: 09.
 Location/Qualifiers
 1..1073

FEATURES
 source

/organism="Triticum aestivum"
 /mol_type="mRNA"
 /db_xref="taxon:4565"
 /clone_id="Triticum aestivum FGAS: Library 5 GATE 7"
 /note="Vector: pCMV.SPORT6; Crown and developmental stages
 of spike formation in wheat cultivar Norstar. 4 mRNA
 populations were combined before constructing the library.
 The first mRNA population is from 1cm crown sections after
 30 days of cold acclimation. The second is from 1cm crown
 sections after 11 days of deacclimation (before
 deacclimation plants were fully vernalized for 49 days).
 The third is from different developmental stages of spike
 formation (5 to 50mm) that still have not emerged from the
 leaf (dissection required). The last is from different
 developmental stages of spike and seed formation after
 having emerged from the leaf (visible). First strand
 synthesis in this library was done in the presence of
 methylated dCMP thereby protecting from internal cleavage
 with NotI."

ORIGIN

Query Match 12.9%; Score 125.2; DB 7; Length 1073;
 Best Local Similarity 49.2%; Pred. No. 2.3e-11;
 Matches 422; Conservative 0; Mismatches 427; Indels 9; Gaps 4;

Db 87 CCGGACGCTGTCGTGCACTCTGTAACGGCCGCCGGAGCGCTGCTGCTGCG 146
 110 CCTCTCTGCGCGCCCGCGCGGAGGCGCGCGCGAAGCTGTCGCCG-GGA 168
 147 CGAGCTACCGCGCGCGCGCGCGGAACTGCGCGCCCTCCGGGACCGCGACTCT 206
 169 TGGCGGCGCGGGCGGATCCCGCCCNCCNCCNCCNCCNCCCGCGCGCGGGCG 228
 207 GCTGCGCGTACACGAGGACGTGCGCTCTCAAGCGCGTGCAGCCGACTCATGCG 266
 229 NCCCCCGCGCGCGNCCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 288
 267 GCGCGGCGCGCTGCTGCGCGACACCTGCTCTGTCGAGCGGGCATGCGCGGAGCTCG 326
 289 GGGGGGGGGGCGCTTCCCGCCNCCCGGAGCCCGCGCGGCGCGCGGGGGCGCG 348
 327 GGGCCAGCGCCCGCGCGCTCCAGACGCGGGCGCTCAACCGATGTTGCGCCCGCGCG 386
 349 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 408
 387 CATACCGCGCGACCGCTGCGCGCGCTGTATACACGAGGAGCGCGCGCGCTCA 446
 409 GCGGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 468
 447 GCTGCGGCTGTCGAGGGCGCGCGCGCGCGCGCGCTGACGCTCACGCGCGAGAGCA 506
 469 CCGCGG---CCCGCGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 522
 507 CCGGACGACGCGCGCGCGACCGAGCGCTGAGGACGCGCGCTCTCTTCCTTCG 566
 523 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 582


```
Query Match      12.9%; Score 124.6; DB 9; Length 1267;
Best Local Similarity 47.1%; Pred. No. 2.8e-11;
Matches 399; Conservative 0; Mismatches 448; Indels 1; Gaps 1;

QY 79 GCGGGACGCGGACGCTGCTGCTGACCTGTACCGCGCGGGAGCGCGGACGCTGCTC 138
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1191 GCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1132
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 139 CTGGTGGGCGACGCTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1131 CGGGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1072
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 CTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 258
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1071 CGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1012
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 259 CTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 318
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1011 TNGCGCGCGG-GGGCGGGCGGGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCG 953
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 GAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 378
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 952 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 893
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 GCGCGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 438
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 892 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 833
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 439 AGCGCCCTGCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 498
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 832 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 773
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 499 GAGACGACGCGGACGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 558
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 772 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 713
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 559 GGGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 618
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 712 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 653
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 619 CCGCACGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 678
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 652 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 593
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 679 GGGGACATCCAGCGGCTGCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 592 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 739 CTGCGCTCTTGTGCGCGCGCTGCTGCGCGGACCGCGGACCGGTGCGACGCGCC 798
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 532 MCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 473
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 799 GCGGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 858
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 472 GCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 413
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 859 GGGGAACTCCGCGCGGCTCATGAGACGCGGAGCTGCGCGCGCGCGCGCGCGAG 918
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 412 CGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 353
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 919 CTGTTCCG 926
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 352 CCGGCGCG 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
AG363333/c      2332 bp      DNA      linear      GSS 03-JUN-2004
LOCUS          Mus musculus molossinus DNA, clone:MSMg01-165A22.TJ, genomic survey
DEFINITION
ACCESSION      AG363333
VERSION        AG363333.1      GI:47974538
KEYWORDS       GSS.
```

```
SOURCE
ORGANISM       Mus musculus molossinus
                Mus musculus molossinus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS       Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE         BAC end Sequences of Library MSMg01
JOURNAL       Unpublished
REFERENCE     2 (bases 1 to 2332)
AUTHORS       Hattori,M., Toyoda,A., Noguchi,H., Kojima,T. and Sakaki,Y.
TITLE         Direct Submision
JOURNAL       Submitted (17-NOV-2003) Masahira Hattori, The Institute of Physical
                and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
                1-7-22 Suenho-chou,Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
                (E-mail:hattori@gsc.riken.jp, URL:http://hgp.gsc.riken.go.jp/,
                Tel:81-45-503-9111, Fax:81-45-503-9170)
                Clones are derived from the mouse BAC library MSMg01. For BAC
                library availability, please contact Kunita Abe (abe@rtc.riken.jp).
                Tsukuba Institute, Bio Resource Center,
                The Institute of Physical and Chemical Research (RIKEN) 3-1-1
                Koyadai, Tsukuba, 305-0074 Japan
                phone: 81-298-36-9189, fax: 81-298-36-9199
                e-mail: abe@rtc.riken.jp
PRIMERS
Sequencing : TJ
LIBRARY
Vector      : PBAC3.6
R.Site 1    : EcoRI
R.Site 2    : EcoRI.

FEATURES
source
1..2332
    /organism="Mus musculus molossinus"
    /mol_type="genomic DNA"
    /sub_species="molossinus"
    /db_xref="taxon:57486"
    /clone="MSMg01-165A22.TJ"
    /sex="male"
    /tissue_type="mixture of kidney and spleen"
    /clone_lib="MSMg01 Mouse Male BAC Library"

ORIGIN
Query Match      12.8%; Score 124.4; DB 9; Length 2332;
Best Local Similarity 48.8%; Pred. No. 2.8e-11;
Matches 421; Conservative 0; Mismatches 436; Indels 5; Gaps 3;

QY 83 GAGCGCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 142
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2163 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 143 TGGGCGACGCTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2103 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2044
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 203 TCTGCTGCGCGCTGACGAGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 259
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2043 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1984
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 260 TCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 319
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1993 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1924
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 320 AGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 379
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1923 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1864
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 380 CCGCGCGCATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 439
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1863 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1804
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 440 CGGCGCTGCTGCGCGCTGCTGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 499
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1803 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1744
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
QY 500 AGACAGACCGGACGACCGGCGCCACCCAGCCCTGACGACGCGGTGCTCTCTCTTCG 559
Db 1743 GCGCGGATCGCGCCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1684
QY 560 GAGTGGCCCTTGGCCGCTCGGCGTCCAGTCCGGGCTTGGCGGCGACGCGCGCG 619
Db 1683 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1625
QY 620 CCGACCGAGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 679
Db 1624 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1565
QY 680 GGGACATTCAGCGGATCCAGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 739
Db 1564 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1505
QY 740 TGGCTCTCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 799
Db 1504 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1446
QY 800 CCGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 859
Db 1445 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1386
QY 860 GGGAACTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 919
Db 1385 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1326
QY 920 TGTTCGCGACCGCTTCACCGCAC 941
Db 1325 GGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1304
```

Search completed: October 5, 2005, 09:01:18
Job time : 3671.67 secs

This Page Blank (uspto)

PC C12R1:645)
CC Transformant producing PF1022 substance and process for CC
CC producing the same
CC and novel biosynthetic gene
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
 1..969
 /organism="Streptomyces venezuelae"
 /mol_type="genomic DNA"
 /db_xref="taxon:54571"

ORIGIN

Alignment Scores:

Pred. No.:	2e-62	Length:	969
Score:	1635.00	Matches:	322
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	6	Gaps:	0

US-10-089-514-6 (1-322) x BD178315 (1-969)

QY 1 MetSerGlyPheProArgSerValValGlyGlySerGlyValValGlyGlyMetPhe 20
DB 1 ATGAGCGGCTTCCCGGACGCTGCTGCGCGGCGAGCGGGGCGGCGGCGCATGTTTC 60
QY 21 AlaGlyLeuLeuArgGlyValGlySerArgThyLeuValValAspLeuValProProPro 40
DB 61 GCGGGGCTGCTGGGAGGCGGGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 120
QY 41 GlyArgProAspAlaCysLeuValGlyAspValThyAlaProGlyProGlyLeuValAla 60
DB 121 GAGCGGCGGAGCGGCTGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 180
QY 61 AlaLeuArgAspAlaAspLeuValLeuLeuValValHisGlyAspValAlaLeuValAla 80
DB 181 GCCCTCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 240
QY 81 ValAlaProValThyArgLeuMetArgProGlyAlaLeuLeuValAlaAspThyLeuSerVal 100
DB 241 GTGGCGGCTGTCACCGGAGCTCATGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 300
QY 101 ArgThyGlyMetAlaAlaGlyLeuValAlaHisAlaProGlyValGlnHisValGlyLeu 120
DB 301 CGAGCGGAGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 360
QY 121 AsnProMetPheAlaProAlaAlaGlyMetThyGlyArgProValAlaAlaValAlaThy 140
DB 361 AACCCATGTTCCCGGCGGCGGAGCTATGACCGGCGGAGCGGAGCGGAGCGGAGCGG 420
QY 141 ArgAspGlyProGlyValThyAlaLeuLeuArgLeuValGlnGlyGlyArgProPro 160
DB 421 AGGAGCGGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 480
QY 161 ValArgLeuThyAlaGlnGlnHisAspArgThyThyAlaAlaThyGlnAlaLeuThyHis 180
DB 481 GTACGGCTCACCGGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 540
QY 181 AlaValLeuLeuSerPheGlyLeuValAlaArgLeuGlyValAspValAlaGlyAlaLeu 200
DB 541 GCCGTCTCTCTCTCTCGGAGCTGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 600
QY 201 AlaAlaThyAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValAlaGly 220
DB 601 GCGGCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 660
QY 221 GlySerProGlnValTyrglyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArg 240
DB 661 GCGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 720
QY 241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260

DB 721 CGGCGGCTGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 780
QY 261 AlaAspAlaProGlyValArgAlaAspAlaProGlyValHisGlyGlySerGlyValGly 280
DB 781 GCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 840
QY 281 AsnLeuAspGlyValAlaPheGlyGlyLeuLeuArgArgLeuMetGlyProGlyLeuValAlaGly 300
DB 841 AACCTGAGCGGAGCTTCTCGGAGGAGCTCGGAGGAGCTCATGAGACCGGAGCTCGGAGGAG 900
QY 301 GlnAspHisCysGlnGlyLeuLeuPheArgThyLeuHisAlaArgThyAspAspGlyGlyVal 320
DB 901 CAGGACCACTGCGAGGAGCTGTTCCGACCTTCACCGGAGCGGAGCGGAGCGGAGGAG 960
QY 321 AspArg 322
DB 961 GACCGA 966

RESULT 2
BD093916
LOCUS
DEFINITION Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes.
ACCESSION BD093916.1 GI:22639504
VERSION BD093916.1
KEYWORDS WO 0123542-A/3.
SOURCE
ORGANISM Streptomyces venezuelae
Bacteria; Actinobacteriae; Actinobacteridae; Actinomycetales; Streptomyces; Streptomyces; Streptomyces; Streptomyces.
REFERENCE 1 (bases 1 to 969)
AUTHORS Yanai,K., Okakura,K., Yasuda,S., Watanabe,M., Miyamoto,K., Mido,N. and Murakami,T.
TITLE Transformants that produce secondary metabolites modified by a functional group(s) and novel biosynthesis genes
JOURNAL MEIJI SEIKA KAISHA LTD, KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO, NAKOKI MIDO, TAKESHI MURAKAMI
COMMENT OS Streptomyces venezuelae
PN WO 0123542-A/3
PD 05-APR-2001
PF 29-SEP-2000 WO 2000JP006783
PR 29-SEP-1999 JP 99P 276314
PI KOJI YANAI, KAORU OKAKURA, SHOHEI YASUDA, MANABU WATANABE, KOICHI MIYAMOTO.
PI NAKOKI MIDO, TAKESHI MURAKAMI
PC C12N15/09, C12N5/10, C12P21/02, C07K11/00// (C12P21/02, C12R1:645)
FH Key Location/Qualifiers
FT CDS Location/Qualifiers
 1..969
 /organism="Streptomyces venezuelae"
 /mol_type="genomic DNA"
 /db_xref="taxon:54571"

ORIGIN

Alignment Scores:

Pred. No.:	2e-62	Length:	969
Score:	1635.00 <td>Matches:</td> <td>322</td>	Matches:	322
Percent Similarity:	100.00% <td>Conservative:</td> <td>0</td>	Conservative:	0
Best Local Similarity:	100.00% <td>Mismatches:</td> <td>0</td>	Mismatches:	0
Query Match:	100.00% <td>Indels:</td> <td>0</td>	Indels:	0
DB:	6	Gaps:	0

US-10-089-514-6 (1-322) x BD093916 (1-969)

QY 1 MetSerGlyPheProArgSerValValGlyGlySerGlyValValGlyGlyMetPhe 20
DB 1 ATGAGCGGCTTCCCGGACGCTGCTGCGCGGCGAGCGGGGCGGCGGCGCATGTTTC 60
QY 21 AlaGlyLeuLeuArgGlyValGlySerArgThyLeuValValAspLeuValProProPro 40


```

/db_xref="GI:47846868"
/translation="MTEQNDLQRLRAELDALDGLTDLTIVRRRIDLVRIARYKSRHGV
PMQGRVSLVKDRAARAYADHGLDESLVNLVYLITTEMCREDVLWMSRESLTADR
R"
gene      /gene="ORF1V"
          4065..5243
CDS       /gene="ORF1V"
          4065..5243
          /codon_start=1
          /evidence=not experimental
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="BAD21143.1"
          /db_xref="GI:47846869"
          /translation="MAATTTAALTGRTAAQCPYPMPDPDAVTEBELAGLPLYTPED
          EVDLNMGMALVARGAALLVGGDCAERFHEAVPDVRRRVNDLQGLAIPRGSLP
          AVAIGRIAGQVGRSSPFTEESGAKRKYCGCAVNVNPEEPARTPERRLLTA
          YTCGQIVDEIRSNMGRPLYERYTSHLELLPYRPLVREAGHGTSGSAHFHMG
          ERTREBDGAVHAAQAVHNPVGVKLPVSBDAVALSRNLNPEGVGRLLTVRGA
          KEVDLLPVPVAVARHGAPEVWLCDPMHNGKLGHKRLLEPMRAETAVFTLR
          EHGMPPAGLHLELTPDVTCEVSELDPRPFRTDRTCDPRLNPEQSAWMVTFLSL
          "
ORIGIN
Alignment Scores:
Pred. No.:      9,25e-62      Length:      5251
Score:          1635.00      Matches:      322
Percent Similarity: 100.00%      Conservative: 0
Best Local Similarity: 100.00%      Mismatches: 0
Query Match:    100.00%      Indels:      0
DB:             1           Gaps:          0

US-10-089-514-6 (1-322) x AB116234 (1-5251)

QY      1 MetSerGlyPheProAspSerValValGlyGlySerGlyValValGlyGlyMetPhe 20
        |||||
DB      2708 ATAGCGCGCTTCCCGCGACGTCGTGTCGCGCGCGCGCGCGCGCGCATGTC 2767

QY      21 AlaGlyLeuLeuArgGluAlaGlySerArgThrLeuValValAspLeuValProProPro 40
        |||||
DB      2768 GCCGGGCTGCTCGCGGACGGCGGCGGCGGACCGCTCTCCATCCGTCGCGCGCG 2827

QY      41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAla 60
        |||||
DB      2828 GGCACGGCGGACCGCTGCTGTCGGCGGACGTACCCGCGGGGCGCGGAACCTCGCGGCC 2887

QY      61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValSA 80
        |||||
DB      2888 GCCCTCCGGGACGGGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2947

QY      81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal 100
        |||||
DB      2948 GTGGCGCGCTGACCGCGCTCATGCGCGCGCGCGCGCTGCTGCTGCGGACACCTGTCG 3007

QY      101 ArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu 120
        |||||
DB      3008 CGACACGGGCAATGGCGCGGAGCTCGCGGCGGACCGCGCGCGCTGCGGACGTCGGGCTC 3067

QY      121 AspProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThr 140
        |||||
DB      3068 AACCCGATGTTGCGCGCGCGCGGATACCGCGCGGACCGCTGCGCGCGCTGCTGCTCAC 3127

QY      141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgPro 160
        |||||
DB      3128 AGGGAACGGGCGCGCGCGCTGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3187

QY      161 ValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
        |||||
DB      3188 GTACGGCTCACGGGCGGAGACGACGACGCGGCGGACCGCGGCTGACGCGAC 3247

QY      181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu 200
        |||||
DB      3248 GCCGTGCTCTCTCTGCGGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCTG 3307
```

```

QY      201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220
        |||||
DB      3308 GCGGCGACCGGACCGCGCGCGCGCGCGCGCGCGCGCGCGCTGCTGCGCTGCTGCTGCTGCT 3367

QY      221 GlySerProGluValTyTrGlyAspLysGlnArgSerAspProArgAlaAlaSerAlaArg 240
        |||||
DB      3368 GGCAGCGCCCGAGGTGTACCGGGAACATCCAGCGGTCCAAACCGCGCGCGCTGCGCGGCC 3427

QY      241 ArgAlaLeuAlaGluValaLeuArgSerPheAlaAlaLeuValGlyValAspAspProAspArg 260
        |||||
DB      3428 CGGCGCTGCGGAGCGCTGCGGCTCTTGGCGCGCTGCTGCGGCGACGACCGGACCGT 3487

QY      261 AlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly 280
        |||||
DB      3488 GCGGACCGCCCGGCGGCGCGCGCGCGCGCGCGCGCGCGCATCCCGGGGGAATGCGACCGCGCG 3547

QY      281 AsnLeuAspGlyValPheGlyGluLeuArgLeuMetGlyProGluLeuAlaAlaGly 300
        |||||
DB      3548 AACCTCGACGCGCTCTTCCGGGGAACCTCCCGGCTCATGCGGACCGGAGCTCGCGCGGCG 3607

QY      301 GlnAspHisCysGlnGluLeuPheArgThrLeuHisArgThrAspAspGluGlyGlyVal 320
        |||||
DB      3608 CAGACCACTGCGGAGAGCTGTTCCGACCTTCACCGGACGACGACGAGGCGAGAG 3667

QY      321 AspArg 322
        |||||
DB      3668 GACCGA 3673

RESULT 4
LOCUS    AF262220
DEFINITION Streptomyces venezuelae chloramphenicol biosynthetic gene cluster,
partial sequence.
ACCESSION AF262220
VERSION   AF262220.2
KEYWORDS  GI:14290414
SOURCE    Streptomyces venezuelae
           Streptomyces venezuelae
           Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
           Streptomycetaceae; Streptomyces.
REFERENCE 1 (bases 1 to 14159)
           He, U., Magarvey, N., Pirae, M. and Vining, L. C.
           The gene cluster for chloramphenicol biosynthesis in Streptomyces
           venezuelae ISP5230 includes novel shikimate pathway homologues and
           a monomodular non-ribosomal peptide synthetase gene
           Microbiology 147 (Pt 10), 2817-2829 (2001)
JOURNAL   Microbiology 147 (Pt 10), 2817-2829 (2001)
MEDLINE   21461106
PUBMED    11577160
REFERENCE 2 (bases 1 to 14159)
           He, U., Magarvey, N. A. and Vining, L. C.
           Direct Submission
           Submitted (01-MAY-2000) Biology, Dalhousie University, 1355 Oxford
           St, Halifax, NS B3H 4U1, Canada
           3 (bases 1 to 14159)
           He, U., Magarvey, N. A. and Vining, L. C.
           Direct Submission
           Submitted (05-JUN-2001) Biology, Dalhousie University, 1355 Oxford
           St, Halifax, NS B3H 4U1, Canada
REMARK    Sequence update by submitter
COMMENT   On Jun 5, 2001 this sequence version replaced gi:10716943.
FEATURES
         source
         1..14159
         /organism="Streptomyces venezuelae"
         /mol_type="genomic DNA"
         /strain="ISP5230"
         /db_xref="taxon:54571"
         /map="adjacent to pabAB"
         /function="essential for chloramphenicol biosynthesis"
         /codon_start=1
         /transl_table=11
         /product="CmlG"
```

CDS

/protein_id="AAC21974.1"
/db_xref="GI:10716944"
/translation="MTAGSPGPGRTPLRPRALAVRDIHSLADIQMDQVNRCTAV
GTYPRGVEDPAPAMREYVARKGLTRALLADIVAEEDVVTLDPRPVGPRLEAD
LVTAHNSRRLRIIDEAVRLA YWRHTDEYVLFVYFSLIYDLADVDLLQIRAHNLG
DTPGRIGSRVRRHARTVEQYTLPAADPAVEVLPALDELVPGRKGLPTISASVEVL
PLEGVEFDELLASVTGLVLPALGGGLVLYQYRSMFARFGYVYIKELVVRGPG
SDYTPVEAKTRHAEISLGRFTMSDLAPARAFTSGMRPVVSDTTFMMPEDRMW
VPRSAVFPEDIKFLADRSWPGPPLLMQYKRGFLSGAALADVDLGLKQIRGASGGTP
"
3498..6404
/note="non-ribosomal peptide synthetase; contains
reductase-like domains near C-terminus"
/codon_start=1
/transl_table=11
/product="CmlH"
/protein_id="AAQ21975.2"
/db_xref="GI:14327897"
/translation="MELPDLITDLEFRHGRGRTALRTAGRTWTEELDRTVSALARID
AECPAGRRVIVAGEHTAEAVVWMLAARSHAVHTPMNPGIPADRFEEPRVADAALIV
CFERREALVRSERKAGRLIYAGDVGMPTDPAAPADGADPAPASRYASIFTSSTGD
PKLVDSHGGLNLCSRLRLDITPDQVLFASISFDASVEILGTLYAGATLVV
VRDQASWLGVSRLAAGCDLAMLSPSVARLDEARSRIKVEFCGALGGEYDX
AARYSRVFNAYGPEATVCFSLAELTSTPSIGTPVDFPAAVADPDSGDHATGSE
LVIVGDVALGYAGSPAEENVEGTVDGPAVATGVVSLSDGELTYLRIDEOIKR
LGRHVNLAHVGSRLSHLGRVLRDPAITLVTADGATEESLMARIQDLPVME
APDRVLVVDALPLTSGKVRSAALRELLASPAAGHGTGDEDAELRLVDVTVAVL
GOEIGPETSIFDAGSSSLAMIQIOVKLSIDYGEBAVAAANADYDPAAPAFRLHNG
EAVAPASAVDTLLRVTETRDRLRAELPLLRDTRIEPVGAADGREGVLLGASGF
IGHVLDRLLAAGRPVIVSTGDDGVLTGAATRGQADYVRAVRSYAELEWRWD
RRGPPVDAVHCGGVNHLPLDLSHSGSVNTALVRAAALGASPAFLSAASAG
ADRLPLSAATLTAVGDPYRSKLISEYVNTLAVGCAVSHYRGLYGRPEDRHIL
KDMFTLTLTARVGMFRLSGHVPICDVTADTLGLGRPDANPGVGLDASRTPDSK
SAVVNHTVYALDELRLRHTGLTEADVLA.PAMFERVGRGVPAFLAAMALSGPM
PSAHRVEDHDLILRLGLTPTDPTAGDRPERTGTTSRPERRRPADTALYGTSP"
3576..3596
/note="Region: adenylation domain A1"
3708..3746
/note="Region: adenylation domain A2"
3539..3586
/note="Region: adenylation domain A3"
4089..4100
/note="Region: adenylation domain A4"
4350..4370
/note="Region: adenylation domain A5"
4494..4538
/note="Region: adenylation domain A6"
4593..4610
/note="Region: adenylation domain A7"
4647..4706
/note="Region: adenylation domain A8"
4833..4853
/note="Region: adenylation domain A9"
4893..4910
/note="Region: adenylation domain A10"
5043..5075
/note="Region: thiolation domain"
5343..5375
/note="Region: NADH binding site"

ORIGIN

Alignment Scores:
Pred. No.: 2,38e-59 Length: 14159
Score: 1588.00 Matches: 315
Percent Similarity: 96.65% Conservative: 2
Best Local Similarity: 96.04% Mismatches: 5
Query Match: 97.13% Indels: 6
DB: 1 Gaps: 1

US-10-089-514-6 (1-322) x AF262220 (1-14159)

QY 1 MetSerGlyPheProAspSerValValIGlyGlySerGlyAlaValIGlyMetPhe 20
|||||

10159 ATGAGCGGCTTCCCGGAGCGTCGTCGTCGGCGGACGGAGCGGCGGCGCATGTC 10218
QY 21 AIGLYLeuLeuArgGLUaIGLYSerAArgThrLeuValValAspLeuValProProPro 40
10219 GCGGGGCTGCTGGGAGGCGGCGACCCGACCCCTGCTGCACCTCGTACCGCGCGC 10278
QY 41 G1AArgProAspAlaCysLeuValIGLYAspValThrAlaProGlyProGlyLeuAlaAla 60
10279 GGAAGGCGGACCGCTGCTGCTGGGCGAGTCAGCCGCGGCGCGGAGCTCGGCGC 10338
QY 61 AlaleuArgAspAlaAspLeuValLeuLeuAlaValHISGLYAspValAlaleuValAla 80
10339 GCCCTCGGAGCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10398
QY 81 ValAlaProValThrArgLeuMetArgProGlyValLeuLeuAlaAspThrLeuSerVal 100
10399 GTGGCGCGCTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10458
QY 101 ArgThrGlyMetAlaAlaIGLYLeuAlaAlaHISAlaProGlyValIGLYValLeu 120
10459 CGGACGGCGTGGCGCGGAGCTCGGCGCCAGCCCGCGCTGCTGCTGCTGCTGCTGCTGCT 10518
QY 121 AspProMetPheAlaProAlaAlaIGLYMetThrGlyArgProValAlaAlaValAlaThr 140
10519 AACCGATGTTGCGCGCGCGCGGATGACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 10578
QY 141 ArgAspGlyProGlyValValThrAlaLeuLeuArgLeuValIGLYGlyGlyArgPro 160
10579 AGGAGCGGCGCGCGCTGACCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10638
QY 161 ValArgLeuThrAlaIGLYLuhISAspAArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
10639 GTAGCGCTACCGCGGAGGAGCGACCGGACGACGCGGCGCGCGCGCGCGCGCGCGCGCGCG 10698
QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu 200
10699 GCCGTATCTCTCTCTCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10758
QY 201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220
10759 GCGGCGACGCGACCGCGCGCCACCGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10818
QY 221 GlySerProGlyValValArgGlyAspLeuGlnArgSerAspProArgAlaAlaSerAlaArg 240
10819 GCGAGCGCGGAGGTGACGAGGACATCCAGCGGTCCAAACCCCGGCGGCGGTCCGCGCGC 10878
QY 241 ArgAlaLeuAlaGlyValLeuArgSerPheAlaAlaLeuValGlyAsp----- 256
10879 CGGCGCTGCGCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 10938
QY 257 -----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274
10939 GCCGAGACCGGAGCGCGCGCCGACGACCGGACCGGACCGGACCGGACCGGACCGGACCGG 10998
QY 275 GLYCysAspGlyValaGlyAsnLeuAspGlyValPheGlyGlyLeuArgArgLeuMetGly 294
10999 GGAATGGAAGCGCGCGGAACTCGAGCGGCTCTCGAGAACTCGCGCGGCTCATGAGGA 11058
QY 295 ProGlyLeuAlaAlaGlyGlnAspHisCysGlnGlyLeuPheArgThrLeuHisArgThr 314
11059 CGGAGCTCGCGCGGCGGCGGACGACGACGCTGCGACGAGCTGTTGCGACCTTCACCGGACCC 11118
QY 315 AspAspGlyGlyGlyValAspArg 322
11119 GACGACGAAGCGGAGAGGACCGA 11142
Db 11059 CGGAGCTCGCGCGGCGGCGGACGACGACGCTGCGACGAGCTGTTGCGACCTTCACCGGACCC 11118
RESULT 5
A48324 A48324 888 bp DNA linear PAT 07-MAR-1997
LOCUS
DEFINITION Sequence 2 from Patent WO9601901.
ACCESSION A48324
VERSION A48324.1 GI:2302117
KEYWORDS

SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
ORIGIN

Streptomyces pristinaespiralis
Streptomyces pristinaespiralis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Streptomycineae; Streptomycetaceae; Streptomyces.
1 (bases 1 to 888)
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,
Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M. and
Dutric-Rosset, G.
STREPTOGRAMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS
Patent: WO 9601901-A 2 25-JAN-1996;
RHONE-POULENC RORER SA (FR)
Other publication AU 2891295 960209
Other publication FR 2722210 960112.
Location/Qualifiers
1. 888
/organism="Streptomyces pristinaespiralis"
/mol_type="unassigned DNA"
/db_xref="taxon:38300"

Alignment Scores:
Pred. No.: 1,27e-20 Length: 888
Score: 661.50 Matches: 155
Percent Similarity: 60.38% Conservative: 34
Best Local Similarity: 49.52% Mismatches: 101
Query Match: 40.46% Indels: 23
DB: 6 Gaps: 4

US-10-089-514-6 (1-322) x A48324 (1-888)

QY 2 SerGlyPheProArgSerValValValGlyGlySerGlyAlaValGlyGlyMetPheAla 21
DB 13 TCGGTGTTCCGGCGCTTGTGTGTGTGGCGGGCGCGTCCGGTGGCCGACATGTTACAG 72
QY 22 GlyLeuLeuArgGluAlaGlySerArgThrLeuValAlaAspLeuValProProProGly 41
DB 73 CACTGTGTGTGGCTGTGGGGGTGGCGGTGACCTGGCTGAC--GTGGCCGGGGCGGT 129
QY 42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla 59
DB 130 GCGGCGGACCGGGGTGGCGGTGTGGCGGTGATGTGGCGGGCGGGCGGGCGGGCGGT 189
QY 60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
DB 190 GCGGCGCTGGCGGGCGGACGTGTGTGTGTGGCGGTGGCGGACCGGTGGCGGTGGAG 249
QY 80 AlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer 99
DB 250 GCGGTGAGGTGTGTGGGGGTGGGTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 309
QY 100 ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly 119
DB 310 GTCAAGACCGCGATCCGCGGGCGGTGCGTGGAGCGGGCGGGCGGTGCGAGGGGTGGG 369
QY 120 LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
DB 370 CTAAACCCGATGTTCCGCCCTCGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 429
QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyArg 159
DB 430 GTACACGACCGGGCGGGGTGTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 489
QY 160 ProValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
DB 490 GTGTGAGAGATGCCGGCGGGCGGACGACGACGTGACCGCGGACGAGCGCGGACG 549
QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
DB 550 CATGCCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 609
QY 200 LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
DB 610 CTGCGGACAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 669

QY 220 GlyGlySerProGluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla 239
DB 670 GCGGCGACCCGAGAGTGTATTTTCAGATCCAGGCGGCCAACCAGCGCGCGCGCGCGG 729
QY 240 ArgAlaGluAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259
DB 730 CCGACGCGCTGTGGCGCGCGCGCTG-----GTCCGCGTGTGGCGGCGCGCTGTGAG 777
QY 260 ArgAlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAla 279
DB 778 AGGCGCGACGAGAGACGTTCCG----- 801
QY 280 GlyAsnLeuAspGlyValPheGlyGlyLeuValArgArgLeuMetGlyProGluLeuAlaAla 299
DB 802 -----GCCCTGTGCCCGAATCGCGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 849
QY 300 GlyGlnAspPheCysGlnGluLeuPheArgThrLeuHis 312
DB 850 CTGAAACGGCTGTGGCGCGGATGTTACCCGCTGCAC 888

RESULT 6
AR198354
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AR198354
Sequence 2 from patent US 6352839.
AR198354
AR198354.1 GI:20248203
Unknown.
Unclassified.
1 (bases 1 to 888)
Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J.,
Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M. and
Dutric-Rosset, G.
STREPTOGRAMINS FOR PREPARING SAME BY MUTASYNTHESIS
Patent: US 6352839-A 2 05-MAR-2002;
Location/Qualifiers
1. 888
/organism="unknown"
/mol_type="unassigned DNA"

Alignment Scores:
Pred. No.: 1,27e-20 Length: 888
Score: 661.50 Matches: 155
Percent Similarity: 60.38% Conservative: 34
Best Local Similarity: 49.52% Mismatches: 101
Query Match: 40.46% Indels: 23
DB: 6 Gaps: 4

US-10-089-514-6 (1-322) x AR198354 (1-888)

QY 2 SerGlyPheProArgSerValValValGlyGlySerGlyAlaValGlyGlyMetPheAla 21
DB 13 TCGGTGTTCCGGCGCTTGTGTGTGTGGCGGGCGGTGCGGTGGCCGACATGTTACAG 72
QY 22 GlyLeuLeuArgGluAlaGlySerArgThrLeuValAlaAspLeuValProProProGly 41
DB 73 CACTGTGTGTGTGTGTGGGGGTGGCGGTGACCTGGCTGAC--GTGGCCGGGGCGGT 129
QY 42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla 59
DB 130 GCGGCGGACCGGGGTGGCGGTGTGGCGGTGATGTGGCGGGCGGGCGGGCGGGCGGT 189
QY 60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
DB 190 GCGGCGCTGGCGGGCGGACGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 249
QY 80 AlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer 99
DB 250 GCGGTGAGAGTGTGTGGGGGTGGGTGGCGGTGTGTGTGTGTGTGTGTGTGTGTGTGT 309

Oy	100	ValaargthnglGmeAalaaGluLeuLeuAalaaHsAalProGlyValGlnHsValGly	113
Oy	310	GTCAAGAGCCGGAGTCCGGGCGCGCTGAGACGGCGCCGAGGCTGCAGCCGGTGGGG	369
Oy	120	LeuAaPromePheAalAProAalaaGlyMetThGlyAArgProValAalAalVal	139
Db	370	CTGAACCCGATTTGGCCCCCTTCGCTGGGTCTTCAGGGGCGGCGGGTGGCGGTG	429
Oy	140	ThraAgaBglyProGlyValThraAlaLeuLeuAargLeuValGluGlyGlyAarg	159
Db	430	GTCAACGACGGGCGCGGTGTGGGGCGCTGTGAGACTGTGTGGCGGGGGGGGGCGG	489
Oy	160	ProValAargLeuThraAlaGluGlnHsAaBArgThrThraAlaThraAlaLeuThr	179
Db	490	GTGTGGAGATGTCGGCGGGCGGACGACGACGTACCGCGCGCGACGACGCGCCAG	549
Oy	180	HisAlaValLeuLeuSerPheGlyLeuAlaLeuAalAargLeuGlyValAaBArgA	199
Db	550	CATGCCCGGTGCTGACCTTCGGAGCTGGGCTGGGTGAGCTGTGGTGAACGTGGGGG	609
Oy	200	LeuAalAaThraAlAProProBProHsGlnValLeuLeuAalAaLeuAalAaGlyLeu	219
Db	610	CTGCGGACAGTGGCCCGCGCGGATCTGCGGAGTGTGGCTGCGCTGGCCGGANTGCC	669
Oy	220	GlyGlySerProGlnValTyrGlyAaBArgLeginAaBArgSerAaBArgAalAaSerA	239
Db	670	GGCGGAGACCGCGAGGTGATTTCCAGATCCAGGCGCCACCCGCGCCGCGCGCGG	729
Oy	240	ArgAArgAlaLeuAalGluAalLeuAargSerPheAalAalAaLeuValGlyAaBArgProAaB	259
Db	730	CGGCAAGGCTGTGGCGCGCGCGCTG-----GTGCGGCTGGGGAGCGCTGGAG	777
Oy	260	ArgAlaAaBArgAProGlyAArgAalAaBArgAProGlyHsAProGlyGlyCyAaBArgA	279
Db	778	AGGGCGCAGAGGACGTTGCC-----801	
Oy	280	GlyAaBLeuAaBArgValAaBheGlyGluLeuAaBArgAaBLeuMetGlyProGlyLeuAalA	299
Db	802	-----GCCGTGTTCCCGCAACTGCGCGGTGTGCTGGCGAGCAGCGCGCGAG	849
Oy	300	GlyGlnAaBArgCyAaGlnGluLeuAaBArgThrGlnHs	312
Db	850	CTGAAACGAGCTGTGCGCGGAGATGTTACCGCGCTGCAC	888
RESULT 7			
A48323/c			
LOCUS	A48323	2888 bp	DNA linear PAT 07-MAR-1997
DEFINITION	Sequence 1 from Patent WO9601901.		
ACCESSION	A48323		
VERSION	A48323.1	GI:2302116	
KEYWORDS	.		
SOURCE	Streptomyces pristinaeaepralis		
ORGANISM	Streptomyces pristinaeaepralis		
REFERENCE	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
AUTHORS	Streptomycineae; Streptomycetaceae; Streptomyces.		
	1 (bases 1 to 2888)		
	Blanc, V., Thibaut, D., Bamae-Jacques, N., Blanche, F., Crouzet, J.,		
	Barriere, J., Debussche, L., Farnachon, A., Paris, J. and		
	Durruc-Rosset, G.		
TITLE	STREPTOKRIMINS AND METHOD FOR PREPARING SAME BY MUTASYNTHESIS		
JOURNAL	Patent: WO 9601901-A 1 25-JAN-1996;		
	RHONE-POULENC RORER SA (FR)		
COMMENT	Other publication AU 2891295 960209		
	Other publication FR 2722210 960112.		
FEATURES			
Source	Location/Qualifiers		
	1..2888		
	/organism="Streptomyces pristinaeaepralis"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:38300"		
ORIGIN			
Alignment Scores:	3.69e-20	Length:	2888
Prod. No.:			

Score:	661.50	Matches:	155
Percent Similarity:	60.38%	Conservative:	34
Best Local Similarity:	49.52%	Mismatches:	101
Query Match:	40.46%	Indels:	23
DB:	6	Gaps:	4

US-10-089-514-6 (1-322) x A48323 (1-2888)

OY	2	SerGIYPheProAgserv[Val]aIGIySerGIyaIaValaIGIyMecPhea1a	21
Db	1824	TGGGTTCGCGCCCTTGTGTGTGTGTGGCGCGGCCGTGCAGTGCGCATGTTACG	1765
OY	22	GIyLeuEuaArg[Val]aGIySerArgThirLeuValaAspLeuValProProGIy	41
Db	1764	CAGTGGCTGTGTGCGCTGGCGGGGTGCGGTGACTGCTGCAGC-----GTGCGCGGGCGCGT	1708
OY	42	ArgProAspAlaCybLeuVal-----GIyAspValThirAlaProGIyProGIuLeuAla	59
Db	1707	GCGCGGAGCGGGGTGCGGGGTGTGGCCGCTGATGTGCGCGCGCGCGGGCCGAGCGGTG	1648
OY	60	AlaAlaLeuAgsApalAAspLeuValLeuLeuAlaValHisGIuAspValAlaLeuLy	79
Db	1647	GCGCGCGCTGGCGCGCGCGCGCACTGTGTGTGCTGGCGGTGCGCGAGCCGCTGCGAG	1588
OY	80	AlaValAlaProvalThirArgLeuMecArgProGIyAlaLeuLeuAlaAspThirLeuSer	99
Db	1587	GCGGTGTGAGCTGTGCGCGGGGTATGTGGGCGCGGTGCGGTGCTGCGGACACTTGTG	1528
OY	100	ValArgThrGIyMecAlaAlaGIuLeuAlaAlaHisAlaProGIyValGIuHisValGIy	119
Db	1527	GTCAAAGCCGGAATCGCGGGCGGCTGTAAGCGCGCGCGGAGCTGCAGCGAGTGGG	1468
OY	120	LeuAsnProMecPheaAlaProAlaAGIyMecThrGIyArgProValAlaAlaVal	139
Db	1467	CTGAACCAGATTCTCGCCCCCTCGCTGTGTTACAGGGCGCGCGGTGGTGGT	1408
OY	140	ThirArgAspGIyProGIyValThirAlaLeuEuaArgLeuValGIuGIyGIyLARG	159
Db	1407	GTCAACGACGGCGCGGTGTGTGGCGCTGTGGAGCTGTGTGGCGGGGTGGGGCGCG	1348
OY	160	ProValArgLeuthirAlaGIuGIuHisAspArgThirThirAlaAlathrGIuAlaLeuthir	179
Db	1347	GTGTGTGAGATCGCGCGCGCGGACGACGACTGACCGCGCGCGCACGAGGCCACG	1288
OY	180	HisAlaValLeuLeuSerPheGIyLeuAlaLeuAlaArgLeuGIyAlaAspValArgAla	199
Db	1287	CATCGCGCGGTCTGGGCTTCTGGGCTGGGCTGGGCTGAGCTGTGTGTGAGCTGTGGGCG	1228
OY	200	LeuAlaAlaThirAlaProProProHisGIyAlaLeuLeuAlaLeuAlaArgValLeu	219
Db	1227	CTGGCGGACAGTGGCCCCGCGCGCACTGGGATGTGGCGCTGTGCGCGGATGCGC	1168
OY	220	GIyGIySerProGIuValTYrGIyAspIleGINArgSerAsnProArgAlaAlaSerAla	239
Db	1167	GCGCGGACGCGCGGAGGTGTATTTTCAGACATCCAGGCCCAACCCGCGCGCGCGCG	1108
OY	240	ArgArgAlaLeuAlaGIuAlaLeuArgSerPheaAlaAlaLeuValGIyAspAspProAsp	259
Db	1107	CGGAGGCGGTGGGCGCGCGCTG-----GTGCGGCTGGGAGCAGCCGTGCAG	1060
OY	260	ArgAlaAspAlaProGIyArgAlaAspAlaProGIyHisProGIyGIyCyAspGIyAla	279
Db	1059	AGGGGCGACGAGAGAGCTTGGC-----	1036
OY	280	GIyAsnLeuAspGIyValPheGIyGIuLeuArgArgLeuMecGIyProGIuLeuAlaAla	299
Db	1035	-----GCCCTGTTCGCGAACCTGCGCGGTGTGTGTGGCGGACGAGCGCGCGAG	988
OY	300	GIyGlnAspHisCySGInGIuLeuPheArgThirLeuHis	312
Db	987	CTGGACCGGCTGTGTGCGCGGAGATTTCACCGCCCTTGAC	949

RESULT 8

AR198353/c	AR198353		2888 bp	DNA	linear	PAT 20-APR-2002
LOCUS	Sequence 1 from patent US 6352839.					
DEFINITION	ARI198353					
ACCESSION	ARI198353					
VERSION	ARI198353.1					
KEYWORDS	GI:20248202					
SOURCE	.					
ORGANISM	Unknown.					
REFERENCE	Unpublished. 1 (baes 1 to 2888)					
AUTHORS	Blanc, V., Thibaut, D., Bamas-Jacques, N., Blanche, F., Crouzet, J., Barriere, J.-C., Debussche, L., Famechon, A., Paris, J.-M. and Durruc-Rosset, G.					
TITLE	Strepocogramins for preparing same by mutasynthesis					
JOURNAL	Patent: US 6352839-A 1 05-MAR-2002;					
FEATURES	Location/Qualifiers					
source	1..2888					
ORIGIN	/organism="unknown"					
	/mol_type="unassigned DNA"					
Alignment Scores:						
Pred. No.:	3.69e-20	Length:	2888			
Score:	661.50	Matches:	155			
Percent Similarity:	60.38%	Conservative:	34			
Best Local Similarity:	49.52%	Mismatches:	101			
Query Match:	Indels:	23				
DB:	Gaps:	4				
US-10-089-514-6 (1-322) x ARI198353 (1-2888)						
OY	2	SerglyPheProArgSerValValAlGlyIserGlyAlaValGlyIsmethPheAla	21			
Db	1824	TCGGTGTCCGGCGCTTGTTGTGTGTGGCGCGGCGCGGTGCAGTTCACG	1765			
OY	22	GlyLeuLeuArgGluAlGlySerArgThrLeuValAspLeuValProProProgly	41			
Db	1764	CACGTGGCTGTGCGCTTCGGGGGTGGCGGTACCTGGCTGAAC--GTGGCCGGGCGCGT	1708			
OY	42	ArgProAspAlaCyLeuVal-----GlyAspValThraAlaProglyProgluLeuAla	59			
Db	1707	GCGGCGGACGGGGGTGCGGTGGTGGCGGTATGTGCCGGCGCGGCGCGGAGCGGTC	1648			
OY	60	AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys	79			
Db	1647	GCGGCGCTGGCGCGCGCGGACGTGTGGTGGCGGTGCCGAGCCGTTGGCGAG	1588			
OY	80	AlaValAlaProValThraArgLeuMetArgProglyAlaLeuLeuAlaAspThrLeuSer	99			
Db	1587	GCGGTGGAGGTGCTGGCGCGGGGTGATGGCGCCCGGTGGCTGCGCGGACACTTGTGC	1528			
OY	100	ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProglyValGlnHisValGly	119			
Db	1527	GTCAAGACCGCATGCGCGCGCGGTGGTAGGGCGCGCGGGCTGCACAGCGGTGGG	1468			
OY	120	LeuSnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAl	139			
Db	1467	CTGAACCCGATGTTGCCCTTCGCTGGGTCTTCAGGGGCGGCGGTGGCGCGGTGGTG	1408			
OY	140	ThraArgAspGlyProglyValThraAlaLeuLeuArgLeuValGluGlyGlyValArg	159			
Db	1407	GTCAACGAGCGGCGCGGTGGTGGCGCGCTGGTAGCTGTGGCGCGGTGGGGGCGCGG	1348			
OY	160	ProValArgLeuThraAlaGluGluHisAspArgThrThraAlaAlaThraAlaLeuThr	179			
Db	1347	GTGGTGGAGATGCCGCGCGCGCGCGCACACAGCTGACCCTGGCGGACAGCGCCACG	1288			
OY	180	HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla	199			
Db	1287	CATGGCGCGGTGCTGCGCTTGGGGCTGGGGGTGAGCTGTGGTGAAGTGGGGGGCG	1228			
OY	200	LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuAlaArgValLeu	219			

Df	1227	CTGGCGGACAGTATGCCCGCCCGCCCATCTGGCAGATCTGGCGGTCTGGCCCGGATGCC	1168
Oy	220	GlycylserProgiuValTYrGlyAspIleGlnArgSerAenProArghAlaISerAla	239
Df	1167	GGCGGAGCGCGGAGGTATTATTTCAGATCCAGGC CGCAACCCCGGCGCGCGCGCG	1108
Oy	240	ArgArghAlaLeuAlaGluAlaLeuArgSerHeAlaAlaLeuValGlyAspAspProasp	259
Df	1107	CGGACGGCGGTGGCGCGCGCTG-----GTGCGGCTGGGCGAGCGCGTCCAG	1060
Oy	260	ArgAlaAspAlaPProGlyArgAlaAspAlaPProGlyHisPProGlyGlyCyAspGlyAla	279
Df	1059	AGGGGCGACAGACGAGACTTGCC-----	1036
Oy	280	GlyAsnLeuAspGlyValPheGlyGlnLeuArgArgLeuMetGlyProGlnLeuAlaAla	299
Df	1035	-----GGCTTTCTGCGGAACCTGGCGGCGTGTGCTGGCGGACGACGCGCGGAG	988
Oy	300	GlyAlaAspHisCYSGInGlnLeuPheArgThrLeuHis	312
Df	987	CTGGAACGGCTGTGCGCGCGGATGTTACCGCCCTCGAC	949
RESULT 9			
SPU60411/c			
LOCUS	SPU60417	4740 bp	DNA linear BCT 07-MAR-1997
DEFINITION	Streptomyces pristinaeaepristalis 4-dimethylamino-L-phenylalanine precursor biosynthesis (papA, papC, papB, papM) genes, complete cds.		
ACCESSION	U60417		
VERSION	U60417.1	GI:1575335	
KEYWORDS	Streptomyces pristinaeaepristalis Streptomyces pristinaeaepristalis Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.		
SOURCE	1 (bases 1 to 4740)		
ORGANISM	Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Zagorec,M., Schleutinger,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and Thibaut,D.		
REFERENCE	Identification and analysis of genes from Streptomyces pristinaeaepristalis encoding enzymes involved in the biosynthesis of the 4-dimethylamino-L-phenylalanine precursor of pristinamycin I MoL. Microbiol. 23 (2), 191-202 (1997)		
AUTHORS	97197164		
TITLE	2 (bases 1 to 4740)		
JOURNAL	Blanc,V., Gil,P., Bamas-Jacques,N., Lorenzon,S., Schleutinger,J., Bisch,D., Blanche,F., Debussche,L., Crouzet,J. and Thibaut,D.		
MEDLINE	Direct Submission		
PUBMED	Submitted (11-JUN-1996) Recherche Pharmaceutique, Rhone-Poulenc Rorer, 13 qual Jules Guesde, B. P. 14, Vitry sur Seine cedex 94403,		
REFERENCE	France		
AUTHORS	Location/Qualifiers		
TITLE	1..4740		
JOURNAL	/organism="Streptomyces pristinaeaepristalis"		
MEDLINE	/mol_type="genomic DNA"		
PUBMED	/db_xref="taxon:38300"		
REFERENCE	68_..2227		
AUTHORS	/gene="papA"		
TITLE	68..2227		
JOURNAL	/gene="papA"		
MEDLINE	/function="D-aminobenzoate synthase"		
PUBMED	/codon_start=1		
REFERENCE	/transl_table=11		
AUTHORS	/product="papA"		
TITLE	/protein_id="AAC44866.1"		
JOURNAL	/db_xref="GI:1575336"		
MEDLINE	/translation="MRTVRLILIDNYSFTYNILOMLAEVNGAAPLVVRNDRTFMOALAPDDPNVVVSPPGPATDTDLGLSRRVITENDLP.LGVCLGHQALCLAGAAYVHAPSEPHORTVDIHHDGGLFRANI PSPLTVRVHSITVROI.PADLRANRHYADQIMAVAHRIURPFYGFHPRESISSSHGRMLANFPDISIRAAQHPRPTERTIPAPAPAPAPAPAPAPAPAPGEYRFLAHVEVAVCVPDADAATPADAPAFWLDSSVEGLAFPTLGAPAPGAGEBITYDVADRARAIVXDGSGETRRBDTLFDHLHEHIAALPATGIPPEFR		

TITLE Complete genome sequence of the entomopathogenic bacterium
JOURNAL Photorhabdus luminescens
REFERENCE Nat. Biotechnol. 11 (1) (2003) In press
AUTHORS 2
TITLE Duchaud, E., Frangeul, L., Ruenick, C. and Kunst, F.
JOURNAL Direct Submission
Submitted (23-APR-2003) L. Frangeul, Institut Pasteur, Genopole, 25
rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail:
lfrangeu@pasteur.fr, fkunst@pasteur.fr
Location/Qualifiers
FEATURES
source
1. .349944
/organism="Photorhabdus luminescens subsp. laumondii T101"
/mol_type="genomic DNA"
/strain="T101"
/db_xref="taxon:243265"
complement (1142. .3038)
/gene="prpB"
/locus_tag="plu3539"
complement (1142. .3028)
/gene="prpB"
/locus_tag="plu3539"
complement (1142. .3028)
/locus_tag="plu3539"
/codon_start=1
/transl_table=11
/product="Propionate--CoA ligase (Propionyl-CoA
synthetase)"
/protein_id="CAE15912.1"
/db_xref="GI:36786847"
/translacion="MSFNAFYQOOSIDPNAFMAEOARIFMHPPEOVLDYNSPPAR
WFGGKTNLCYNALDRESOPAKLITITSTNSEKVTPEKHELOENRRAAMLS
LVKKGDVLYVPMIAEALFYLLACARIGAHVSVFGFASOSLATRLDNAPVLY
SADGARGGKIIPYKPLDBAIELANHRVHVMNRGLAEVWVPGRIDIPATLKQK
YPDADIPTVMBESNETSCVLYTSGTTPKQVQDGYVALATSMDIIFGKAGEV
PFTSDIGWIVGHSYIAYAPLIGMATINIEGILPIPDAGIWMKIVEKQIVRMGAP
TAIRVLYKKYPIECIGKDISSLKTLIAGEPLDEHARWIAEAINVPIVDNWOTFG
WPIMAIARTLDRBPRSGSTGPMVGFHKLINELGEACGDNKMLVIEGPIPGCG
IOTIVGDPTRFINTYWRHREHVAYSTFDMGINSDDPYFILGSDVINVAHGRICR
EIEECTASHEDVAEVAIVGIKAIKQVAVAPVLDGKEIQAEHFALKEVLMGLV
NKQISVGRPARIFYPSQLPKTRSGMLRRTWQALCEGREPGLSLIENPASIVIRE
AVFC"
complement (3033. .3038)
/gene="prpB"
/locus_tag="plu3539"
complement (3101. .4565)
/gene="prpD"
/locus_tag="plu3540"
complement (3101. .4552)
/gene="prpD"
/locus_tag="plu3540"
/codon_start=1
/transl_table=11
/product="2-methylcitrate dehydratase"
/protein_id="CAE15913.1"
/db_xref="GI:36786848"
/translacion="MSTSVASNHREYDVIADIVYVMEYAITSSVAYATAYCPIDT
LCCGALAEYPRACKLIGPIVETVBNVARGPTQFDLPVOAANITMTMRMDEN
DTWLAWEHGPSDNLGGILAVADMVSRNAVAGKPLTIDKVLTGAVKHAIEOGCAL
EKAENKVGADHVVLVYVASTAVAVAEMLGLNREBIIISAVSLAWDQGSLETPEYAPAT
BSKSNVAGDATSAAVLAALMAQGEWGPVSLTAKTGFYDVLFGNGQPKFORYAGY
VWENVLFXIAFPAEFHAOTAVEAATSLHQOLDAGKTINDITKITIRTHEACIRIIDK
OSPIINPADRDHCIOYMAVPLIFGLTGAADYEDSVAAPRIDALREKIVCEDDDFT
RHYHDEKRSINALTIVITDGSOLDSEVVEFFIGARRKEGIPILLKPFNTLARO
PEFTQORILSLVSDLSLESMVEVNYDLVYI"
complement (4560. .4565)
/gene="prpD"
/locus_tag="plu3540"
complement (4607. .5787)
/gene="prpC"
/locus_tag="plu3541"
complement (4607. .5773)
/gene="prpC"
/locus_tag="plu3541"
/codon_start=1
/transl_table=11

RBS
/product="2-methylcitrate synthase (Methylcitrate
synthase) (Citrate synthase 2)"
/protein_id="CAE15914.1"
/db_xref="GI:36786849"
/translacion="MSQONTMTELTTPFKKSVALSQVAGNTALCTVCKTGNDAHYR
GYDIIIDATQCFEEVAYLVHEKLPSSRAELAAVKKALTLALGDPSSVAALETLPAA
AHPMVMTRGVSLGCTLPEDKDNHLAGARDIADRLALSSILLYWHYSHNGRIE
VBTIDVYISGHLHLHNGKCTCESEWKNHTSLIYAEHEFNASTFSTRVIAGTGDI
YAAIGALIGLGRPHKGANEVSFEIQREYNPDAAERDILARLARKEVYIGFGHFDY
TISDPRHAKVIKQVALQLSKEAGTKMFIADRLAEVMMANKKMPFLDWFSAVSYHM
GVPTMFPFLFIARTAGAAHIIIEORIDNKIIRSDVITGPENRVPFTIGRG"
complement (5782. .5787)
/gene="prpC"
/locus_tag="plu3541"
complement (5834. .6736)
/gene="prpB"
/locus_tag="plu3542"
complement (5834. .6724)
/gene="prpB"
/locus_tag="plu3542"
/codon_start=1
/transl_table=11
/product="Probable methylisocitrate lyase
(2-methylisocitrate lyase)"
/protein_id="CAE15915.1"
/db_xref="GI:36786850"
/translacion="MTLSAGLAFFRAIDABSPLOVGTINANHALAKRAGYKAIYL
SGGVSAGSLGMPDICI STLDDVLDIHDITGVCLPLVDIDGSSAFVAVTRV
SMIKAGAGHILIEDOVGAKRCGRNKEIYSEBEVDRIRKAVDARTDGHFTIMARTD
ALAVGPAALADRARAYIAGADMILPEALITLHMVKPASTGTGVPVLANITTEFGATP
LETTBELLSADVSIALYPLSAPRANKAABQYITLRRRGTSKVISMWQTNELYES
INYVPEQGLDILFPOKKSQ"
complement (6731. .6736)
/gene="prpB"
/locus_tag="plu3542"
7047. .6724
/gene="prpA"
/locus_tag="plu3543"
7047. .7052
/gene="prpB"
/locus_tag="plu3543"
7057. .8724
/gene="prpA"
/locus_tag="plu3543"
/codon_start=1
/transl_table=11
/product="Probable catabolism operon regulatory protein"
/protein_id="CAE15916.1"
/db_xref="GI:36786851"
/translacion="MAETPSAERDNRDPRVITWVSRLFDLFRDITPEFDHOANITP
IOLGPNNAVQHIRQRLTTECDVAVCAGSNGAYLKSRLSPVYIAKASGDFPMOLAR
AHOISIRIGVITYONTLPALBEFOORPNLRITRORSYVTEEDBAROVNLKATGIVYI
GAGLITDVAIEAIVGVFVYSDSIRQAFPHLDIARTMTLSQASISAYPTNNRTR
HTLNDIRGDSAMBEHVRNTIMLYASPATVVLQSGSGTKEILVQAALHETALRGQV
AGKHORPFAVAVGCAIASLLEAELFYEBAFTSSRGGAGLGFVAVHGRITLFLDEI
GEMPLQTRRLRVIEBSKVAVGGHRRPIAVDVRITCATHCIDIDTVMQGRPTLDFY
RSLVATRNIPALREBGHIGIALLEYLRAPALNLPVOSOSVCELATCOBALOVHM
PQNVRLNIMERLITLYISAPHEHILTEBLITALSPEOSITLKOAEQSKSSPSSIQQ
SSIVYPLSVHETLKFPSNHNHTAIVLGISRTYLMRLKKEIARERE"
complement (8738. .9709)
/locus_tag="plu3544"
complement (8738. .9709)
/locus_tag="plu3544"
/function="unknown"
/note="unnamed protein product; Similar to probable
transcription regulator"
/transl_table=11
/codon_start=1
/transl_table=11
/protein_id="CAE15917.1"
/db_xref="GI:36786852"
/translacion="WNITKIVFLIFPNVNLDSLGPLOYFTTANFLAKFGHRRPYAI
KVACDPQGVMTQAGITVQAEPLPLDLSPADTLVAGGIVGESSANHPPLRWMLGHA
SLIARIVSGAGFIILACGLIDGRVYVTHWSSCAELARRYSKLRVECDPIFIODGPV

Best Local Similarity: 32.06% Mismatches: 126
Query Match: 23.36% Indels: 38
DB: 6 Gaps: 5

US-10-089-514-6 (1-322) x AK770909 (1-349980)

QY 8 ValValaValaGlySerGlyValaValaGlyValMetPheAlaGlyLeuLeuArgGluAla 27
DB 115977 GTCATCTTGGTGGGCAAGGCGCTATGTGTTCTTTATTAGAGCCCTTTTACTCAATAT 116036
QY 28 GlySerArgThrLeuValValaAspLeuValProProProGlyArg----- 42
DB 116037 GGCATATTGTATACAGTGTGTCAACGACGAGGAGTGGTCTCATTAATCAT 116096
QY 43 ProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeu 62
DB 116097 TCTCATTAAGTGTAGATT---GATATCTTAACCCGCGCTGTATGTGAGCGGTTTG 116153
QY 63 ArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaAla 82
DB 116154 AGTGGGCGAAGGCGGCGCTTTGGCTTACCGGAGCGCTCCCATACAGCGTTGCC 116213
QY 83 ProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr 102
DB 116214 TGGGTCTTATCTGCTATGTGTATGATGATACAGATGCTCAACCTGTTCAAGTTCAAGGCC 116273
QY 103 GlyMetAlaAlaGluLeuValAlaHisAlaProGlyValGlnHisValGlyLeuAspPro 122
DB 116274 CCTTTTATTTCTGTCATTAAGAGCTGCCGCCGCGGAGCGCTTTCATGTGTTAAACCG 116333
QY 123 MetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAsp 142
DB 116334 ATGTTCTCCGCACTTTGGCCGAGCAAAAAGGCCAGTTGCATCTTCGATTAACGT 116393
QY 143 GlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyValArgProValArg 162
DB 116394 CATGTTAGTGAACATTGATGAGCGGTATATGATGAGCGGTATATGCTATCACTAG 116453
QY 163 LeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaVal 182
DB 116454 ATGACCCCGAGAGAACACGATCAGTTATGCGCTTTGTACAGCTTTGCCCATGCCGCG 116513
QY 183 LeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValArgAlaAlaAla 202
DB 116514 ATACTATTTTGGCTTGGTGGCGCAAAAGTCCGTAGATATCGTACCTTGGCAGCT 116573
QY 203 ThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySer 222
DB 116574 ATTGCCCGACCGCCCATGCTACCATGCGTGGCGTTATATCAAGATCTTAACGCAATCCA 116633
QY 223 ProGluValAlaGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgAla 242
DB 116634 ATGGAAGTGTACTGGGACGTTCAATATGAAATCCGATGCTGCCAAGCGCTCAGAG 116693
QY 243 LeuAlaGluAlaLeuArgSerPheAlaAlaLeuVal-----GlyAspAspProAspArg 260
DB 116694 TTCTCAGATCCATACACAGTATGATTCGATAGTACAGGCTGCGCATCAACTGAA--- 116750
QY 261 AlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly 280
DB 116750 ----- 116750
QY 281 AsnLeuAspGlyValPheGly-----GluLeuArgArgLeuMetGlyProGlu 296
DB 116751 -----TTTGGCATTAACCTCAGAGATATTCGCAACACACTGGGGGAACAG 116795
QY 297 LeuAlaAlaGlyGlnAspHisCysGlnGluLeuLeuArgThrLeu 311
DB 116796 CTGTGATATCAGTCTGCTGATTTGCCAGCATATTTTTCACAGCTT 116840
RESULT 12
AE010371
LOCUS AE010371 14192 bp DNA linear BCT 01-AUG-2002

DEFINITION Methanopyrus kandleri AV19 section 70 of 157 of the complete genome.
ACCESSION AE010371 AE009439
VERSION AE010371.1 GI:19887260
KEYWORDS
SOURCE
ORGANISM Methanopyrus kandleri AV19
Methanopyrus kandleri AV19
Archaea; Euryarchaeota; Methanopyri; Methanopyrales;
Methanopyraceae; Methanopyrus.
REFERENCE
AUTHORS Slesarev,A.I., Mezhevaaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozayvkin,S.A.
The Complete Genome of the Hyperthermophile Methanopyrus kandleri AV19 and Monophyly of Archaeal Methanogens
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 14192)
AUTHORS Slesarev,A.I., Mezhevaaya,K.V., Makarova,K.S., Polushin,N.N., Shcherbina,O.V., Shakhova,V.V., Belova,G.I., Aravind,L., Natale,D.A., Rogozin,I.B., Tatusov,R.L., Wolf,Y.I., Stetter,K.O., Malykh,A.G., Koonin,E.V. and Kozayvkin,S.A.
Direct Submission
JOURNAL Submitted (04-FEB-2002) Fidelity Systems, Inc., Gaithersburg, MD 20879
FEATURES
source
Location/Qualifiers
1..14192
/organism="Methanopyrus kandleri AV19"
/mol_type="genomic DNA"
/strain="AV19"
/db_xref="taxon:190192"
71..1321
/gene="MK0793"
71..1321
/gene="MK0793"
/codon_start=1
/transl_table=11
/product="Uncharacterized conserved protein"
/protein_id="AA02007.1"
/db_xref="GI:19887261"
/translation="MPSFLLEGARQLTIEGVAYGKDPHAYVEDAIDQTEPNIQGLV FSPYHIEIESAVGSLPEGCKVTCSTAGELTPDGYYGTIVTALISSPYLAVETHR IIFEREDPQHAREILRNVLGMEKLRNPGCLDINPAMLTSLSEGRIRALPYFTI BLVEGLVPADIDYMDGINEVQKFLVEYGCAAGDDRLERTVYLEGRTLCPEGSAV VAFSGTAKTIGGALRCGFEPEYEDPEPTKAVPERRIYFEGEPAADYAAVVGKPV BELDSDVPMNMNIGWEIRGRIIEVEPEVDEDSMIPHSRPAAGALIRLEPTENM RETARARVERAMRPDIRDPEETALVLPDCAHRCDAQYDAIREVGDVPIVGFKT VGEHQLESQSPVGHFNQITLEVDR"
1337..2524
/gene="MK0794"
1337..2524
/gene="MK0794"
/codon_start=1
/transl_table=11
/product="Uncharacterized protein conserved in archaea"
/protein_id="AA02008.1"
/db_xref="GI:19887262"
/translation="MTSTSVTVVSREITKGVSEIRVGIVIHGPTVIDGMARIRSL KRLGVRKLTGCTTGVALLDGLDSVYVFPDKLPSECLSEINEMADYIVLTNPKSKR ESGLPAFEOVLTSRAGVDSLVQAEKRGEDGGLVLMNPEDARTVAEHUREDLKITL EVRTVYTERGGIGSRVACVEPGRILVNGITVGVASRNVLEVFPDSDGLVIRIG GRIKPEGVRLGVDPERVVYKTDRLRTEREKRVSGPERIHRVILVDDARKV EDMRSDAVSVGDDTTCVCAELGRLGVWVGLVLDLPDGVNRODTERSLASSENLA ALVCERDDDAKLVKRGKFRFRERKRVLDPTVVEIVVEVCEKVKYHHAKE TSA"
2581..3564
/gene="MK0795"
2581..3564
/gene="MK0795"
/codon_start=1
/transl_table=11
/product="Selenophosphate synthetase"
/protein_id="AA02009.1"

[illegible][illegible]

Db 6520 CGCGAGCTC-----GACGTCGAGGCT---GCTCCGACCAACGTCGATCCGCCCA 6566
63 GAAPAAAPLeuValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaPro 83
6567 AGATGCCGACGTTGTGTGTGTTCCGTCGCCCATCTCCGTACTCGAAGACGTCATCCGTGA 6626
Qy 83 oValThrArgLeuMetArgProGluAlaLeuLeuAlaAspThrLeuSerValArgThrGrl 103
6627 GGTTCGACCAACGTTCTCTGAGGGTTCCCTCTCACCAGACGTGACCTCGTCAAGGTTGC 6686
Db 103 yMetAlaAlaGluLeuAlaAlaHisAlaPro---GlyValGlnHisValGlyLeuAsnPr 122
6687 TCCCGTTCGGCTATGTCGAGACACGTCGCCGAGATGTCGTACGTCCTCGGACTCACCC 6746
Qy 122 oMetPheAlaProAlaAla---GlyMetThrGlyArgProValAlaAlaValAlaThr-- 140
6747 CTTCTTCGGCCCAACGTCGCCCATGCTCCCGGCTCAGACCTGATCTTCACGCCCAACGGA 6806
Qy 141 -ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyValArgPr 160
6807 GCGTAGTGGTCCTCGACCTCGCCGCGTTCGTTATCTCGACGGGAAAGCGCTGAGT 6866
Qy 160 oValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHi 180
6867 CGTTGAGACCAACCCCGAGGAGAACACATGTCACCTATGCGCCGTCGTTCACTGCTCACCA 6926
Qy 180 sAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArg-----LeuGlyValAla 196
6927 TCGTGTCTCTCTGCTGTCGACGCCCCCGCATGAGCCGTTCTCTCCCAAGTCTCAGCTCGA 6986
Qy 196 pValArgAlaLeuAlaAlaThrAlaProProProHiGlnValLeuLeuAlaLeuAla 216
6987 CATCGAAGAGCTGCGCCGAC-----CCCGTGTATCGCTCATGAGACCTCGTGGG 7037
Qy 216 aArgValLeuGlyGlySerProGluValTyrglyAspIleGlnArgSerAsnProArgAl 236
7038 TCTATTCGCGGCGGAGATCAAGCTTTTACGCGGATCAAGCCTTCAACCCCTTACG 7097
Qy 236 aAlaSerAlaArgArgAlaLeuAlaGluAlaLeuAlaArgSerPhe 250
7098 TGACGAAGCCCTGTAAGAGCTCTCCGCGCCTCAAGCCTTC 7140
Db

RESULT 13
AE011181 12180 bp DNA linear BCT 03-APR-2002
LOCUS Methanosarcina acetivorans str. C2A, section 526 of 534 of the
DEFINITION complete genome.
ACCESSION AE011181 AE010299
VERSION AE011181.1 GI:19918739
KEYWORDS
SOURCE
ORGANISM
Methanosarcina acetivorans C2A
Methanosarcinae
Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
Methanosarcinaceae; Methanosarcina.
1 (bases 1 to 12180)
Galagan,J.E., Nussbaum,C., Roy,A., Endrizzi,M.G., Macdonald,P.,
Fitzhugh,W., Calvo,S., Engels,R., Smirnov,S., Arnoor,D., Brown,A.,
Allen,N., Naylor,J., Stange-Thomann,N., Dearellano,K., Johnson,R.,
Lincoln,L., McEwan,P., McKernan,K., Talamas,D.E., Grahame,D.A.,
Zimmer,A., Barber,R.D., Cann,I., Graham,J., Tittell,A., Ye,W.,
Guse,A., Hedderich,R., Ingram-Smith,C., Kuetner,C.H.,
Krzyski,J.A., Leigh,J.A., Li,W., Liu,J., Mukhopadhyay,B.,
Reeve,J.N., Smith,K., Springer,T.A., Umayam,L.A., White,O.,
White,R.H., de Macario,E.C., Ferry,J.G., Jarell,K.F., Jing,H.,
Macario,A.J.L., Paulsen,I., Pritchett,M., Sowers,K.R.,
Swanson,R.V., Zinder,S.H., Lander,B., Metcalf,W.W. and Birren,B.
The Genome of *M. acetivorans* Reveals Extensive Metabolic and
Physiological Diversity
Genome Res. 12 (4), 532-542 (2002)
JOURNAL
MEDLINE
PUBMED
11932238
REFERENCE
2 (bases 1 to 12180)

AUTHORS Birren,B.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Center for Genome Research, Whitehead
Institute, Nine Cambridge Center, Cambridge, MA 02141, USA
FEATURES
source Location/Qualifiers
1.12180
/organism="Methanosarcina acetivorans C2A"
/mol_type="genomic DNA"
/strain="C2A"
/db_xref="taxon:188937"
complement(541..4614)
/gene="MA4589"
complement(541..4614)
/gene="MA4589"
/codon_start=1
/transl_table=11
/product="cell surface protein"
/protein_id="AA07928.1"
/db_xref="GI:19918740"
/translation="MNIRSKYAVYIPFLILPFGTASALPTVAHDRVQGEVYAS
TAMDSHYSTNSFDVNGTVFARYVGVVAGTSPSISATFNGKAFETNPAYISGM
GVTVIPEVTVQVQPGTNTATIDSHNGDGRQVGSITLVVLENKPEVEVIAEGL
DMVIGYGSSTYFNDNTIDLDVDSALYSTILITGYSSESLNGESLSGAVDSGAYF
NTRKDNVKSULTAENQITI KVVGDSDYTSVAVHGLSTYIKLPDLVPTSLTBSYVTPGT
VMTKATIENTDKDPSRNVSLIADGTVDQDVAGIAGGSTVVDLQMTPEGKDSY
SLTVVVDPDNSVSEETNNELTVLVSASVEVPANFTVAPTSGLIPLTVFTDOST
NSPTSMFMDPGDGENSTSONPSHTYISGNHVSILVTNAEGDSEVKDYITVEVIT
PVAATNANTSGAAPFTVQPDOSANILPAEMFMDGDCNSTEONPAHTYITAGKYV
KITTNMAAGNDESEKTDYITAYIPVANFTATPTSGGAPLTYTPTDSTNNPTBMLD
FDDGNSVDQNIHTYTSAGNTVSFTATNSGDSSEKTAITVITYAEKGGLADTA
MKPFHYDIGNTGQSPYTGTPQANVVMVITYTTEMLGOGAVIGSDGTWYIIGRSGLTY
AINPGLTKMGTSIGTSKSPPTPAISDGTIYAGNDKIYVAINPDGSIKMSYPTKLS
VSSPAISADGTIYATSTFALPKPDGTLKMSAVSGYVYFSPALGSDGTIYAGT
YCGALSNADGTLTAMTYSIPAPNNYITDTPALASDGTIYREGCNANIALYALPDGTLK
WTVSVCSTASQAKKSPAGSDGTIYRSGNNOYFALYPDGTLKMSKRAKSGCQTPT
TOSDGTITIDNGSKTLYAFNPDGTVKMSYADGNVYASPAISDGLTFKGGYTVTA
FRDVAVSNTFADPTSGAPLTVNFTDSTGSPSWFMDPGANSTQNPAHTYVA
GNYTNNLYTNSAGADPEVKIDYITVLPSSTPEPVAFTADVTSGLAPLTVNFTD
STGSPSTMLMDPGDSTNSTEONPHTYVLSAGSYTNVLTVENDAGSDPELKTIDYINSE
TSGSVTVTYFDPRESSVAENESTETSIIVASNPAGLSGNTLVAILDDPAVAVETIYK
PTWALITERSITPGTSYIMKYTYLDEDSIKEEGEVVLATUTYSGKEKSNANISIVK
LEDDSGDSLBPALITGLTILBVTLLSLPDEYAPKLDGDLGYEDLTNGEFSFVDIVA
YFHNMDWI EENNPPVEYFDFNGNGRIDFVDVDMFAMI"
complement(5828..6397)
/gene="guaa"
/note="synonym: MA4590"
complement(5828..6397)
/gene="guaa"
/codon_start=1
/transl_table=11
/product="GMP synthase (glutamine-hydrolyzing)"
/protein_id="AA07929.1"
/db_xref="GI:19918741"
/translation="MRELKILVNNVYGGFCHLIRAVBDLMDTKIIPNVPTIEDILA
EHPDGLISGGEEMERAGICPDYAEIDIPILGICQALALAGCHVSKKKGGA
BIEIIVIEDDLIRGLGPKITVMAHAEVALIPGFIHLASDICEIEMAHSPKKPI
YGVQWHPVESHKKGDDELTNFEVCDBY"
6911..7708
/gene="deoc"
/note="synonym: MA4591"
6911..7708
/gene="deoc"
/codon_start=1
/transl_table=11
/product="deoxyribose-phosphate aldolase"
/protein_id="AA07930.1"
/db_xref="GI:19918742"
/translation="MSEIGKRIIRLNRNSRNWYIIPMDGISDGPTEGILNITDT
VNRVVEGGANAVIMQKGMVYKGRGVGIDIGIIVHISVSPBNKAKVOVCTVEV
IKMGADVSMHINVSNTNEDDLBGLKISDCTCKWNPILAMVPRKKITINPDPY
NVAHARIAGELGADVKTIVTIGDPSFRDVRCGPVPVVIAGGFKSTIDELLEMID
GAIEAGARGAALGRNVFOHRDPRVLRITRAICETIVHRRPVEALEQLK"
7816..8958

gene

/gene="MA4592"
 7816..8958
 /gene="MA4592"
 /codon_start=1
 /transl_table=1
 /product="conserved hypothetical protein"
 /protein_id="AA07931.1"
 /db_xref="GI:19918743"
 /db_xref="GI:19918743"
 /translation="MKKSWIKADGEGMEQKDRITTGESGADCVLPNGVEKRR
 ELGNTVAARNDKSRADIVVKGREGGQTKPLQEIQSGFDINATLLMDKGVV
 GGVYVTKDKHYEHFAEMGKICDYLVTGDMKVIPLNLIADLQHKVITIGVSA
 BEARLAFQTELAGDGLDLSGNPQEIKDITIKARLESESALEAAVTRVPLMG
 DRVCVDTCNLMGEGMLIGSOAGMFLVNSESDSPDYVAAPRRVAVAGVHSYIKG
 EKTXYLSEIADGPRTIVDSKGRREGIVGRVKTIESRPLMLEKADRLTALQNA
 ETIKLVGKDGTPTSVALEKDEVLVLEBARHFGKITEITIEK"
 9118..9846
 /gene="arod"
 /note="synonym: MA4593"
 9118..9846
 /gene="arod"
 /codon_start=1
 /transl_table=1
 /product="3-dehydroquinase dehydratase"
 /protein_id="AA07932.1"
 /db_xref="GI:19918744"
 /db_xref="GI:19918744"
 /translation="MTQIGPDELEKAAVAVVILEKPLENSKRAEKAGDILVRLDL
 LGIRNPSAKITREIKSETGLPVLTRNSVABGKGEKVDTEILLVLSKQSP
 DAVIDELASREDRDVYKAKAGKTVIISSHDFSTPSPQENTATLAMEFLAEDI
 AKIAPMGSMEDVNLKVLLEFKNIGTKVCTIAMGRKHTRVAVLQSGVLTVASI
 ESNVAVAPQLPDEVKIMEMLK"
 9843..10685
 /gene="arod"
 /note="synonym: MA4594"
 9843..10685
 /gene="arod"
 /codon_start=1
 /transl_table=1
 /product="shikimate 5-dehydrogenase"
 /protein_id="AA07933.1"
 /db_xref="GI:19918745"
 /db_xref="GI:19918745"
 /translation="MKRVFGVFGDPVGHSLSPAMHNSAFALGMDCIYHAFRPEPL
 RKALIGEAAMGFGNLTVLKEBALKLDLFRPDLBERIGAVNTVVFSEGEIRGN
 TDLGARGOALLBAVEIRGSKVVAVAGGABAVALADAGAEITVNTTEBAVTL
 AKDYAASLPGKINGTSGLKELRADLLINTTIGMHPNTDTTIAIBELHSGLT
 VFDIYNPLETRLKEAKVAGAKTVSGVLMVYQGAAPFLMTGVEAPALMKTVILE
 ALQA"
 10682..12112
 /gene="MA4595"
 10682..12112
 /gene="MA4595"
 /codon_start=1
 /transl_table=1
 /product="prephenate dehydrogenase"
 /protein_id="AA07934.1"
 /db_xref="GI:19918746"
 /db_xref="GI:19918746"
 /translation="WNPDPKPSGSKTKVLLGSGMGQFTRFFKERYKYTTWCK
 GGVAVARLNVPPASLDLAIPENDIVIVSPINVEETIAEVAAPMKAGSLMDPT
 STRKVPVEMARAPAGVEILGTHMGFPPIPIRGQTVLIPVKESEKMPVIRDL
 FEEGAAVEITTAABHDRLSVVQGLTHFAVISGTTIDRLDFIKSRKFPVSI
 MLDPVGRILGONPLYLAIQENPGIPEVHAPEECEEISLVRADESGFVRKMA
 AARKYDDTALALRRSDKLINSRIAEVETLLNSVKGCGFSHVSGRPHGVKQGN
 EVILTKASGTSPIHINKRTIKLKLNTLRLSESELEMERREKLVHSVRDISVLIPG
 ADPEIVAGAVTDBHLVACQISDIYDGETERTEDKNEIRETRGLGVYRITTFDCNAD
 SVBAEVTALLCGLCGRVREKRLKLE"

US-10-089-514-6 (1-322) x AE011181 (1-12180)
 QY 2 SerGlyPheProArgSerValValValGlySerGlyValAlaValGlyMetPheAla 21
 Db 10709 TCCGAAAAAACAGAGTTTGTGATCTTGGCGGAAACGAGATGCGGAGTGGTACT 10768
 QY 22 GlyLeuLeuArgGluAlaGlySerArgThrLeuVal-----ValAsp 35
 Db 10769 CGTTTATTTTAAAGAAAGGGTTTAAAGTTACCGGTCTGGGAAAAGCGGCAAGTGAA 10828
 QY 36 LeuValProProProGlyArgProAspAlaCysLeuValGlyAspValThrAlaProGly 55
 Db 10829 GTGCCGAGAACTGATGATGATCCCTTTGCT----- 10858
 QY 56 ProGluLeuAlaAlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAsp 75
 Db 10859 TCGGAACCTTACCCAGCATCCCTGAAACGATATAGTATGATTCCTCCGATCAAC 10918
 QY 76 ValAlaLeuValAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAla 95
 Db 10919 GTAAACGAAAGACCATGCAAGTCGCCCCGAGAGTAAACAGGAGACCTTCATG 10978
 QY 96 AspThrLeuSerValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaPro---Gly 114
 Db 10979 GATTTTACCTCCACCAAGTAAAGCTGTGAAGCCATGACAGAGTTTGACCTGCCGG 11038
 QY 115 ValGlnHisValGlyLeuAspPheAlaProAla---AlaGlyMetThrGlyArg 133
 Db 11039 GTTGAAGATTCGGAACGACCCCATGTTGGGCCCCAGATCCCACTCAGGAGCAAA 11098
 QY 134 ProVal-----AlaAlaValValThrArgAspGlyProGlyValThrAlaLeuLeuArg 151
 Db 11099 ACGGTATCTCTTGTCCCGTAAAGAGCGCTCGGAAAATGTTCCGTAATCAGGACG 11158
 QY 152 LeuValGluGlyGlyGlyGlyArgProValArgLeuThrAlaGluGlnHisAspArgThr 171
 Db 11159 CTTTGTGAGGAAAGCGGAGCGCCACGTTGAAATACAGACCGCAACGACGACGCTG 11218
 QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191
 Db 11219 GTCTCGGTACTACGAGGCGCTGACGCAATTGTTGTTATATCTCAATVAGAGACCATTTGAC 11278
 QY 192 ArgLeuGlyValAspValAlaArgAlaLeuAlaAlaThrAlaProProProHisGlnValLeu 211
 Db 11279 AGCCTTGATTTCCAGATCAAGAAATCCAGGAAATTCGTAAGTCCGCTACAGATATG 11338
 QY 212 LeuAlaLeuLeuAlaArgValLeuGlyGlySerProGluValThrGlyAspIleGlnArg 231
 Db 11339 CTGACTTGTGGGCAAGATCTCTGGCCAGAACCCCTTACCTTACCGCCCTGATCCGATG 11398
 QY 232 SerAspProArgAlaAlaAlaSerAlaArgAlaLeuAlaGluAlaLeuArgSerPheAla 251
 Db 11399 GAAACCCCGGATATTCGAGAGTGAAGCATGATTTATAGAGAAATGTGAGAACTTCC 11458
 QY 252 AlaLeuVal 254
 Db 11459 AGCCTGGTA 11467
 RESULT 14
 LOCUS AE013358
 DEFINITION Methanosarcina mazei strain Goel, section 140 of 379 of the complete genome.
 ACCESSION AE013358 AE008384
 VERSION AE013358.1 GI:20905730
 KEYWORDS
 SOURCE Methanosarcina mazei Goel
 ORGANISM Methanosarcina mazei Goel
 ARCHAEA; Euryarchaeota; Methanomicrobacteria; Methanosarcinales; Methanosarcinaceae; Methanosarcina.
 REFERENCE 1 (bases 1 to 9888)
 Depgenmeier,U., Johann,A., Hartech,T., Merkl,R., Schmitz,R.A.,

ORIGIN
 Alignment Scores:
 Pred. No.: 0.00748 Length: 12180
 Score: 272.00 Matches: 76
 Percent Similarity: 45.63% Conservative: 44
 Best Local Similarity: 28.90% Mismatches: 123
 Query Match: 16.64% Indels: 20
 DB: 1 Gaps: 5

TITLE
JOURNAL
MEDLINE
PUBMED
22120827
12125824
2 (bases 1 to 9888)
AUTHORS
REFERENCE
Deppenmeier U., Johann, A., Hartesch, T., Merkl, R., Schmitz, R. A.,
Martinez-Arias, R., Henne, A., Wierer, A., Baumeier, S., Jacobl, C.,
Brueggemann, H., Lienard, T., Christmann, A., Boemecke, M., Steckel, S.,
Bhattacharyya, A., Lykidis, A., Overbeek, R., Klenk, H.-P.,
Gunsalus, R. P., Fritz, H.-J. and Gotschalk, G.
The genome of *Methanosarcina mazei*: evidence for lateral gene
transfer between bacteria and archaea
J. Mol. Microbiol. Biotechnol. 4 (4), 453-461 (2002)
TITLE
JOURNAL
Submitted (23-AUG-2001) Goettingen Genomics Laboratory, Institute
for Microbiology and Genetics, Grisebachstrasse 8, Goettingen
37077, Germany
FEATURES
source
1. .9888
/organism="Methanosarcina mazei Gol"
/mol_type="genomic DNA"
/strain="Gol"
/db_xref="taxon:192952"
242. .817
/gene="MM1269"
242. .817
/gene="MM1269"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAM30965.1"
/db_xref="GI:20905731"
/translation="MKRRHRIATISDKHWLKLKIHAKERTQKALEIALECIENNT
OSELTVQKRYLMNCESIGSVCCQVQKALIMETVNHSEFKYVTKNPTYEIOL
LQKPLNCSLEEVNGLMIVFRTSHMFDTIDCKDNDGYLLILTHSLGKNSKLNMT
CKSLFTTGGADISKTSGNIIFIKVFKDKTL"
complement (1025. .1594)
/gene="MM1270"
complement (1025. .1594)
/gene="MM1270"
/EC_number="6.3.5.2"
/EC_number="6.3.5.2"
/codon_start=1
/transl_table=11
/product="GMP synthase [glutamine-hydrolyzing]"
/protein_id="AAM30966.1"
/db_xref="GI:20905732"
/translation="MRRLKILVNNYQGFCHLIHRTVRDIMDKIIPNVPIEIIA
EERPDGILSGSPMERAGLCFDYVREIDVPIILGICGHQIALAYGSHVAGKGGYA
EIEVEVLBEDDILRGLGPKTTVASHADEVALIPDGFILHARSVDCEIEMMRPTKPI
YGVQMHPEVSHTKKGEILMNFPEVCOLY"
2056. .2859
/gene="MM1271"
2056. .2859
/gene="MM1271"
/EC_number="4.1.2.15"
/EC_number="4.1.2.15"
/codon_start=1
/transl_table=11
/product="2-dehydro-3-deoxyphosphoheptanote aldolase"
/protein_id="AAM30967.1"
/db_xref="GI:20905733"
/translation="MTMSEIGKKIRIERLNRRESRMVVIPLDHGISDGPBEGILNIT
DTYNRVABEGANAVLMOKGMVRYGHRGYGHIDGLVYHISASVLSPPRAKVOCTVE
EYIKGADVASHAINVSGDTEDALEMLGKISDCTCKWMPILAMNYPRGKKTINPHD
PVNVAAHARIAGELGADVKTIVTGPDSFRDVRGCPVPVYLAGSPKSTJDMLELM
IDGAMGAGARAGIARNVFOHDPVRLTRAICEIVHRRRVEALEQLK"
3000. .4142
/gene="MM1272"
3000. .4142
/gene="MM1272"
/EC_number="4.6.1.3"
/codon_start=1

/transl_table=11
/product="3-dehydroquinase synthase"
/protein_id="AAM30968.1"
/db_xref="GI:20905734"
/translation="MKKSVWIKABEGGNEBOKERTITTYESGACDVLNPPDVENR
ELGNITVAAPFARDNKSADIVVYGRKSGDTRPPEIIFSGFDVNAATLIDKGTIV
GGYVVIKDRSYRFAAEMKICDYLLVTGTDWKAIVPLENLIDLQREKVIIFGVSA
EBARLAFQTLFETGADVGLDSCNPOEIKDTIKARALESESTELSAVTVREPLGMG
DRCVDTCNLMORGEMLIGSOAGMFLVNSGDSBPYVAAFPVYNAGAVSHYKIG
DKTRYLSELQAGDSVTITDSRGKREGIVGRIRIKISRPMLILEARAGDRTLVAIQNA
ETIKLVKGGGTISVAKLKKGDEVILRLEBGRHKKIEITLIK"
4203. .4931
/gene="MM1273"
4203. .4931
/gene="MM1273"
/EC_number="4.2.1.10"
/codon_start=1
/transl_table=11
/product="3-dehydroquinase dehydratase"
/protein_id="AAM30969.1"
/db_xref="GI:20905735"
/translation="MTQIGSFPLEKKTAVNAVILRKPLENSKKAABKAGDILFBLD
LCIRTPERAAYIRREIKATGTPIIVTSSGABGGKMDCKEDRGLNLINLSKDP
DAIDIELSAGMKERKNKVIKAADRETAIVSSHDPKTPLELQNMRTIIEEMPLAGDI
AKLAVPLPSVGDNLNLRVTLDPKDAKGSVCTIANGSGDKTRVVAAPFGSVLTVASI
ESNASAPGQLPVDKIKIMMLK"
4928. .5770
/gene="MM1274"
4928. .5770
/gene="MM1274"
/EC_number="1.1.1.25"
/codon_start=1
/transl_table=11
/product="Shikimate 5-dehydrogenase"
/protein_id="AAM30970.1"
/db_xref="GI:20905736"
/translation="MKQVGVGDPVGHSLSPAMNAAFSALGMDCIYHAFVPTBEKL
EKAILGAEMGFGGINTLVPLKEKALKDFVPDPLAKRIGAVNTLVFEKKODIGYN
TGPGKQKALDPTGVEIRGSRMVVAGAGAAVAFOLAADGADITVIRTEBRLVQ
ABEISAADLPGKIRGTGSLGKELRDADVILNTTIGMHPTDATTIVAEELHSGLT
VVDIYVNPLETTLRLBARTSGAKTISGVLMLVYQGAEARLMTGVBPPLMKKTYLE
ALQA"
5767. .7185
/gene="MM1275"
5767. .7185
/gene="MM1275"
/EC_number="1.3.1.12"
/codon_start=1
/transl_table=11
/product="Prephenate dehydrogenase"
/protein_id="AAM30971.1"
/db_xref="GI:20905737"
/translation="MNTDYDAKTKVLLIGTGEMGQWTRPFKRGVBYTWGKGGK
IEVAKLDVPFALDLEAVIPESDIIIVSVPINATEETIAELAPKKAKGSLINDFTSIK
VGPVEAMRKFAKPDVEIILGTHPMFGPTTIRINGQTVIIPVGVSEKMPVRLQFEE
SGAHVEITTAEBHDLVSVQGLTHFAVYAIQTTIDRLDPVKSRKFPVYSINLD
FYGRILGONPVYALIOENRPGVPVHEAFICEKELSLVIAHMBEGFVRMKAAR
KYDDTSHARRSDKILNSRIIEYETILNSTGVCGFSHTYSGNHYGRLEKRPHEIV
LMKLVSKGTAPIIKRFTTLKLENLRPUSEALREKRENLSHVRSVIVPEGDH
EVLRAVSTNKKHLADCKISDIYGVNGTLLERKGSYAEKGLVYRISIFGDADDSVE
TEVTALLCGLGRIREKMLKFS"
7260. .8795
/gene="MM1276"
7260. .8795
/gene="MM1276"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAM30972.1"
/db_xref="GI:20905738"
/translation="MSQVKKWISLLFALITTSIFPSLLISDKFTESNGENVNTAMA
GVSGENOVISLNNRLJETNGSEARLNAATELGNLGEPAASAIIEKMEENPELGEIK
SYMLLALLETGDKRAENIISVYLKKEASNAADVQSGKEVPEDIMLAIOAKKAM

RESLARSMTSEYNDITDLEALKEBONSTLYTSFVLSNFSESENETILLOALKS
EKGVYVSAAMALGOKKEKAAATPILKMLADODPLAGHSAVMAVLMGEMDRANVTM
ELKONKDYRSSTALALGLKEBAEAVPILERLPRASVRENLSVLGKMDENAV
BPLKNTLESKDESKERKOSLNTGVDRKSTVLAIGIGTSTQTLIGVINDGEIP
DVAVVAATIALGNIGSPEAVNLKTYENOMNNVNRGALIALGKTENODAAEFIK
IGKDDGVSAAREALINNGEFAVEPLIENLKTEDOKIKDETALILIEIGDPRAVEPLIE
AYK"

complement (8942. .9691)
/gene="MM1277"
complement (8942. .9691)
/gene="MM1277"
/EC number="2.1.1.-"
/codon_start=1
/transl_table=11
/product="methyltransferase"
/protein_id="AA030973.1"
/db_xref="GI:20905739"
/translation="MMETKEIIRNYMDHRESEIYSGIVESEERLAKMKLAKLGG
RKLEILVDGTGQALAMFAEMGHHLTAVDLSAMLEKARKNALORSLDINFIGDA
ENLOPDHFDVVSRRKILWTLDPDKALLEMRVLRKKDMITAIIDGWYSGIFLKS
IRTVSGIIRSVKERNRNDPFRKNTQULKNDLPLYSLKRGVFGLLDAGFEENIERM
DSLCLSRKRGSLDKLDYSHPIYFIRAVK"

ORIGIN

Alignment Scores:

Prod. No.: 0.0301 Length: 9888
Score: 256.00 Matches: 71
Percent Similarity: 46.77% Conservative: 52
Best Local Similarity: 27.00% Mismatches: 119
Query Match: 15.66% Indels: 21
DB: 1 Gaps: 5

US-10-089-514-6 (1-322) x AB013358 (1-9888)

QY 2 SerGlyPheProArgSerValValValGlyGlySerGlyValValGlyGlyMetPheAla 21
DB 5785 GCTGGAAAACGAAAGCTTCGATCTTGGCGGACCGAGAGATGGGCGAGTGTTCACA 5844
QY 22 GlyLeuLeuArgGlyValGlySerArgThrLeuVal-----ValAsp 35
DB 5845 CGTTTTTTCAGAAAGAGGCTATGAGCTTACTGTGGGGAAGGCGGCAAGATGAA 5904
QY 36 LeuValProProProGlyArgProAspAlaCysLeuValGlyAspValThrAlaProGly 55
DB 5905 GTGCGCAAAAACCTGCGACGTAACCTTTTGCTTTA-----5937
QY 56 ProGlyLeuValAlaAlaLeuArgAspAlaAspLeuValLeuValAlaHisGlyAsp 75
DB 5938 ---GACCTTAGGAGATTATCCCTGAGAAAGTACATAGTATGTTCCGTCGATCAAT 5994
QY 76 ValAlaLeuValAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAla 95
DB 5995 GCAACAGAAAGAAACATCGCAGAAATCGCCCAAAATGAAAGCTGGAACATCTTATG 6054
QY 96 AspThrLeuSerValArgThrGlyMetAlaGlyLeuValAlaHisAlaPro---Gly 114
DB 6055 GATTATTCCTCAATAAAGGTAGACCTGTGAACCGTGAAGAAAGTTTGCGCCCAAGAT 6114
QY 115 ValGlnHisValGlyLeuAsnProMetPheAlaProAla---AlaGlyMetThrGlyArg 133
DB 6115 GTTGAATTCCTTGGAGCAGACCCCATGTTTGTCGACATTCACACATAAGGGGCA 6174
QY 134 ProValAlaAlaValValThrArg-----AspGlyProGlyValThrAlaLeuLeuArg 151
DB 6175 ACAGTATTCCTGTTCCGTAAGAGGCTATTCGGAATAATGTTTCCGTAATCAGACAG 6234
QY 152 LeuValGlnGlyGlyGlyArgProValArgLeuThrAlaGlnGlyHisAspArgThr 171
DB 6235 CTTTTTAGGAAACGGCGCGCACGTTGAGATCAACAATCGAGAGAACGACGAGGCGTG 6294
QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191
DB 6295 GTCTCAGTACGACGGGCGCTTACTCATTTTGCTTATGCTATAGCCACACATTCAGC 6354

QY 192 ArgLeuGlyValAlaAspValArgAlaLeuAlaAlaThrAlaProProHisGlnValLeu 211
DB 6355 AGGCTGATTTTACGTCACAAAAGTCACAGAAATTCGTAAAGCCGCTTTACATATATG 6414
QY 212 LeuAlaLeuLeuValaArgValLeuGlyGlySerProGlyValValGlyAspIleGlnArg 231
DB 6415 CTCGACTTGTGGCGAGATCTCGGACAGAAATCCATCTCATATGCCCGATTCAGATG 6474
QY 232 SerAsnProArgAlaAlaAlaSerAlaArgAlaAla-LeuAlaGlyAlaLeuArgSerPheAl 251
DB 6475 GAAACCTCGAGTCTCTGAGTGCATGAAGCATTTATAAAGAGTGTGAGAGCTTCC 6534
QY 251 aaAlaLeu 253
DB 6535 AGCCTTG 6541

RESULT 15
BX957223
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS

TITLE

JOURNAL

FUNDED

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

gene

CDS

1.268122
/organism="Methanococcus maripaludis S2"
/mol_type="genomic DNA"
/strain="S2"
/db_xref="taxon:267377"
227..1555
/gene="secY"
/locus_tag="MMP1422"
227..1555
/gene="secY"
/locus_tag="MMP1422"
227..1555
/locus_tag="MMP1422"
/evidence=not experimental
/transl_table=11
/product="Aldehyde dehydrogenase:SecY protein"
/protein_id="CAF30978.1"
/db_xref="GI:45047860"
/db_xref="UniProt/TREMBL:O6LKD0"
/translation="MESFLIKIKPLIELIPYKRPRLRISPKKLOMTALVIVLPIL
GTTDITGSSNPALRFDFTQYTTASMGTLITLIGCTPTTAAAGTINOLVGSBLISDL
SKPMNALRQGLQKLFGLACPLDALMFVAGAPFALVTLMTAVLVFQALGAILITY
LDEIVSRVIGSIGIGLFIAGVSQITFVGTFCABEGVLMKFTFAMTASGLMTALVILP
ILGTTILVPLVVVYVESIRVEIRPLAGRVGAVGAKYIKPIYVSNLPSVILAAALFANIO
LMGMFLKKGFPLIGHYTGKAVDGLAAYFSTPVGISSTADPLHAFVFTVMVIVICI
LRLFLVETNSGIDAKSMARKLNDMAINGPRKSGISIRQLRKRYITPTVWGSARVG

misc_feature
FLAAADFTGALGGTGVLLTVTSIVRYLVEQLVQQLSEIHPAIKIRK"
order(317, .385,443, .511,572, .640,650, .709,728, .796,
854, .922,995, .1063,1181, .1249,1352, .1420,1430, .1489)
/gene="secy"
/locus_tag="MMP1422"
/note="10 probable transmembrane helices predicted by
TMHMM2.0"
complement(1595, .2494)
/locus_tag="MMP1423"
complement(1595, .2494)
/locus_tag="MMP1423"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="putative glycosyl transferase, family 4:Aldehyde
dehydrogenase"
/protein_id="CAF30979.1"
/db_xref="GI:45047861"
/db_xref="UniProt/TREMBL:O6LXC9"
/translation="MDLYEVLALPCFVLSLFTKEMIKKMYKYGYDLHKADKIKV
AEWGLSPVSVVMLEFNPALSLIFPGFVGVIDISRLSKKIKVLTLPFV
APFKLGLPSSLIIIGIFVSSNLTMNLGPNGLETGMGILCLFMAVCIAMGIDFG
FKLLIFLAAYLGLLYNNRPAYFPFGDTGTEPIGAFLATIAYRGRPIBELPLMAP
MDALRKQFTAGVTKDSVFTPTQLKNGKLCVGGYLSLPRMLMKKAMEEKIVLVL
MALEAFFGLISLYTRKFCFNIP"
complement(order(1613, .1678,1853, .1918,1964, .2020,
2042, .2107,2153, .2218,2255, .2320,2423, .2479))
/locus_tag="MMP1423"
/note="7 probable transmembrane helices predicted by
TMHMM2.0"
2591, .5149
/locus_tag="MMP1424"
2591, .5149
/locus_tag="MMP1424"
/BC_number="2...-"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Oligosaccharyl transferase, STT3 subunit"
/protein_id="CAF30980.1"
/db_xref="GI:45047862"
/db_xref="UniProt/TREMBL:O6LXC8"
/translation="MGEFLNKVSDPFKKKEKIKIILLFTIGMMSFOIRAOQADMAFT
DNSYLODMFDDNGMYLTALDPYYLRMTENYVNDYSVGETVIGENIPYTI
OYAPGRBAGLVALSLATVLYSVWNSIDSTVTIMNAAFVPAIMSIPLGIPVEFIV
RRNTASNGIGLVALLLISSPLLTKTSGAFSDPTPEILPILFIYMMIMEAHBEN
SKKSGIFGGIAAILIGIYPMMSGMYAEDITRGPVLVLTAYEYLTKSKLKVITTS
LITLVGAILVLSLSTGLSGFINWILSPIGTVINEAKITGMNVNMYSELAIPTVT
DIENSVGNIMWLSLIGISLISFVFKDKQKIDIKVLYLTMIAIAYVATAKIR
FVALMTPALAIGIGIYFAGQIENIKRYEKKEVILYVIGISLVITLIKYGGLAFL
VPTTVPPIAYVLSIALLVLAAYKIIDISEKEQAKVKGIGILLAFMLPPSMAALVP
FYTAPTMNNGMDSLSMISKSTPENSVCVMONGHITMAYATPKMTPTPGGSCNTPRA
YWGAFSTSDENLSVGIIRMLATSGDSAYDDSIILKTGSIKQTVDILINKLIPLTR
TEAKASLVNNYDILDAEAEVLDLTHPKVTNPYLLITNNMTSIAVSWMFGWNSL
PASTENSDRMGYYQOLGSAODINGTIVVYIPLQETDSRVINILEITDSELSKNSA
VIDSNQTSWQSPNFHKLILKNGVNYEQTENNGDYSIEIVREKLSDEGYOYAAVWS
SKULEDSIYTKLHFLDGYGLKESISEKESVDPRSYGIQPGFKVYSVDYGRDYLN"
2591, .2701
/locus_tag="MMP1424"
/note="Signal predicted by SignalP 2.0 HMM (Signal peptide
probability 0.992) with cleavage site probability 0.966 at
residue 37"
order(2648, .2707,2999, .3067,3086, .3145,3173, .3232,
3257, .3310,3320, .3379,3413, .3481,3611, .3670,3704, .3754,
3767, .3826,3863, .3922,3950, .4018,4055, .4123)
/locus_tag="MMP1424"
/note="13 probable transmembrane helices predicted by
TMHMM2.0"
5262, .5801
/locus_tag="MMP1425"
5262, .5801
/locus_tag="MMP1425"
/codon_start=1

/evidence=not_experimental
/transl_table=11
/product="protein of unknown function DUF127"
/protein_id="CAF30981.1"
/db_xref="GI:45047863"
/db_xref="UniProt/TREMBL:O6LXC7"
/translation="MAIELRLGHRGEBDKRISTHVALTSRALGAEKIIFTEEDKHVK
ESVERIVEAMGGEFKEVYKSWRSYAKPKDNGIIVHLLTMGENINKIMTEIRIDISK
TKNLLIITIGAERVPREAYDLADVNLISVGNQHSVLAALAIRLDFTECKTLYSEYED
AKIKVTPSKSEKCVPEKD"
5825, .6445
/gene="dcd"
/locus_tag="MMP1426"
/locus_tag="MMP1426"
/note="synonym: duc"
5825, .6445
/gene="dcd"
/locus_tag="MMP1426"
/BC_number="3,5,4,13"
/BC_number="3,6,1,23"
/note="Citation: J. Biol. Chem. 278:11100-11106(2003)"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Bifunctional dCTP deaminase/dUTP diphosphatase"
/protein_id="CAF30982.1"
/db_xref="GI:45047864"
/db_xref="UniProt/TREMBL:O6LXC6"
/translation="MILSDKDIENVYTSKRVLEPNSKPGSCSYDTLGGSEFIKYN
EDVYDVKTLNKSRTKRENSVAVCCPLNHLDDTTIKNKKKKNVDCVSGSLGTTNE
YVELPNDVCAQYQGRSSFGFVPLQTHQAGWIDSGFKITLIEIAYADFPVILYNQR
VQQLFSLKTLSPADIGYSDRKCSKYAGQTSVWASDIKDFETNEBE"
6449, .7825
/locus_tag="MMP1427"
6449, .7825
/locus_tag="MMP1427"
/locus_tag="MMP1427"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="Conserved Hypothetical Protein"
/protein_id="CAF30983.1"
/db_xref="GI:45047865"
/db_xref="UniProt/TREMBL:O6LXC5"
/translation="MFVPEEIIETIKMIEYQMLDIRTTTIGINLKCADKQDLIKEN
YDKTSLGCLNIVETANKSOKYGIPIVKRISTVPIGLIMSTYKGLSDEBAVACV
EKGITLDAKRGVDFGTSYALVOKRATYEKMLISIPMLMTKTDKVCASVAVAT
TKAGINMYVKKMGRIIVKETSSETITDAGCAKIYFCNAPBNPMAAGFHPGSGDA
VTVAGSGPVRAVVEQLKQKQIGVSDIEIKTIAFKITRMGELVKEVANSLVNFG
IYDLSLAPPAIGDSIANLLEAVNGERCGTHGTALMLNDVAKKGMASSNVGL
SGAFLPVSDPACMIABAVEGALRLKLEAMTCVCGGLDMLIVPEKTPASTLSAAMD
EMAGIMINKKTTAAVNIIPVPGKDVGDVYEGGLGTAPIMPVSFSSBELIBRGRIIP
APIQSLIN"
7898, .8059
/locus_tag="MMP1428"
7898, .8059
/locus_tag="MMP1428"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="hypothetical protein"
/protein_id="CAF30984.1"
/db_xref="GI:45047866"
/db_xref="UniProt/TREMBL:O6LXC4"
/translation="NAKNQDAGLSLSAGLVRYMDDVSKIKIAPKVLGLTVSIIIE
AYLVANGILIT"
7994, .8053
/locus_tag="MMP1428"
/note="1 probable transmembrane helix predicted by
TMHMM2.0"
complement(8114, .8431)
/gene="rpm"
/locus_tag="MMP1429"
complement(8114, .8431)
/gene="rpm"

Alignment Scores:

Pred. No.:	3.24	Length:	268122
Score:	239.00	Matches:	67
Percent Similarity:	47.47%	Conservative:	55
Best Local Similarity:	26.07%	Mismatches:	127
Query Match:	14.62%	Indels:	8
DB:	1	Gaps:	5

US-10-089-514-6 (1-322) x BK957223 (1-268122)

```

QY      9 ValValGlyGlySerGlyValValGlyGlyMetPheAlaGlyLeuArgGluAlaGly 28
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80133 ATATCGCGCGAAGCTGCGGCTGGAATAATGTTGCATCATTTCTTAATAATATAGGGGT 80192

QY      29 SerArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeuVal 48
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80193 TACGATGTACTCGTATCTGGAAGGATGTATATAAGAAAAGMAGTTGAAAAAGAACTT 80252

QY      49 GlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArgAspAlaAspLeuVal 68
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80253 GGT---GTAACTTACACAAATAACAAATATCGATGCTCA---GAAAAAGAGATATTGTA 80306

QY      69 LeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeuMet 88
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80307 ATATTTCTGTCTCCCTATTAATGTTACAGAACTGTGTATTAAGAGATTGCTCCGCATGTT 80366

QY      89 ArgProGlyValAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaAlaGluLeu 108
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80367 AAAAAAGGTTCACTTTTATATGACATTAACATCGATCAATAAGAAAAGGCTCGCAAAACGATG 80426

QY      109 AlaAlaHisAla---ProGlyValGlnHisValGlyLeuAsnProMetPheAlaProAla 127
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80427 GCAGAAATTTCAAAAGCTGAGATATTGTCTTCGCCACACCCCTATGTTGGGCCAGAA 80486

QY      128 AlaGlyMetThrGlyArgProValAlaAlaValAla-----ThrArgAspGly 143
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80487 ACTCCTTCGTTAAACCGGACAGTTGTAACTTAACCTACTGAAAAAGAAAAACAAT 80546

QY      144 ProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyGlyArgProValArgLeu 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80547 CCATTTTGTGAAAAAATTAAGAATTTTGAATAATGMAAGTCAAAAGTAATTTGTTGTA 80606

QY      164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80607 TCCCAAAAGAACACGATTAATTAATGCGTGTAGTTCAAGGCTCACTTTTGTATAT 80666

QY      184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80667 ATTCACTTGATCCCACTCACTTAAGAATCTTGGAATAGATATAAAGAAATCCAGAACTTT 80726

QY      204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerPro 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80727 GCGTCCCAATTTACGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 80786

QY      224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgAlaLeu 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80787 TACCTATATGACATATCCAGATGACACACCCCAATATAAAACATTTCAGAAACATTTC 80846

QY      244 AlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 80847 ATTAATAAACTGCAAAATATTAAGTGAATTTGTC---CAGATTAAGGATAGG 80894

```

Search completed: October 5, 2005, 15:36:38
 Job time : 5131.02 secs

This Page Blank (uspio)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 4, 2005, 21:38:23 ; Search time 565.456 Seconds
(without alignments)
3371.006 Million cell updates/sec

Title: US-10-089-514-6
Perfect score: 1635
Sequence: 1 MSGFRSVYVGGSGAVGGMF.....HCOELFRTIHRDDEGKDR 322

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4390206 seqs, 2959870667 residues
Total number of hits satisfying chosen parameters: 8780412

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+pn.model -DEV=xlh
-O/cgn2.1/USPTO.spool.h/US10089514/funat.04102005.105743.8027/app.query.fasta_1.1621
-DB=N Geneseg -QMT=fastcap -SUFFIX=ring -MINMATCH=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blodsum62 -TRANS=human40.cdt -LIST=45
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcio -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514 @CGN 1.1 860 @runat.04102005.105743.8027 -ICPU=6 -ICPU=3
-NO WMAP -LARGEOUTERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -MARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
1: N Geneseg_16Dec04:*
2: geneseqn1980s:*
3: geneseqn1990s:*
4: geneseqn2000s:*
5: geneseqn2001as:*
6: geneseqn2001bs:*
7: geneseqn2002as:*
8: geneseqn2002bs:*
9: geneseqn2003as:*
10: geneseqn2003bs:*
11: geneseqn2003cs:*
12: geneseqn2004as:*
13: geneseqn2004bs:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1635	100.0	969	4	AAAF6411 4-amino-4
2	1635	100.0	969	8	AAAL50182 S venezue
3	1588	97.1	3305	8	ABZ69799 Plasmid p
4	1588	97.1	12391	8	ABZ69798 Plasmid p
5	661.5	40.5	888	2	AAT59269 Streptomy

C	6	661.5	40.5	2888	2	AAT59268	AAT59268 Streptomy
	7	382	23.4	888	10	ACF71774	ACF71774 Photorhab
	8	382	23.4	110000	10	ACF67367_49	Continuation (50 o
	9	382	23.4	110000	10	ACF65387_1	Continuation (2 of
	10	323	19.8	110000	11	ADM27081_07	Continuation (8 of
	11	237	14.5	1122	8	AAD50753	Aad50753 Escherich
	12	235	14.4	110000	2	AAV21209_09	Continuation (10 o
	13	223	14.3	1122	6	ABN83448	ABN83448 Escherich
	14	223	14.3	1238	6	ABN99612	ABN99612 E. coli c
	15	223	14.3	1238	6	AA146032	AA146032 E. coli ty
	16	227	13.9	1131	11	ACH98777	ACH98777 Klebsiell
	17	225.5	13.8	1122	8	AAD50752	Aad50752 Erwina h
	18	225.5	13.8	1122	12	ADJ98178	ADJ98178 Pantoea a
	19	225.5	13.8	1152	6	ABN83449	ABN83449 Erwina h
	20	224	13.7	3957	13	ADT05414	ADT05414 Haemophil
	21	224	13.7	349980	13	ADT05737	ADT05737 Haemophil
	22	216	13.2	110000	2	AAT42063_13	Continuation (14 o
	23	213	13.0	1792	5	AA65801	AA65801 DNA encod
	24	200	12.2	1125	10	ADP00853	ADP00853 Bacteri
	25	177.5	10.9	1122	10	ACF69143	ACF69143 Photorhab
	26	177.5	10.9	110000	10	ACF67367_20	Continuation (21 o
	27	177.5	10.9	243072	10	ACF65382	ACF65382 Photorhab
	28	165	10.1	2025	13	AD661947	AD661947 Bacteri
	29	163.5	10.0	110000	4	AA199682_39	Continuation (40 o
	30	159	9.7	88421	6	AA140781	AA140781 88421nt g
	31	152	9.3	891	13	ADR31472	ADR31472 Chori
	32	150.5	9.2	1362	8	ABX56041	ABX56041 M. echino
	33	150.5	9.2	36602	13	ADR01210	ADR01210 Farnesyl
	34	150	9.2	42291	9	ACC58253	ACC58253 Chlorobac
	35	149	9.1	1236	9	ADA337164	ADA337164 Human SUV
	36	149	9.1	2732	2	AAT43625	AAT43625 Chromatin
	37	149	9.1	2732	12	ADP10568	ADP10568 Reference
	38	149	9.1	71138	12	AD013830	AD013830 dbv gene
	39	148	9.1	1023	13	ADR41634	ADR41634 Lipid acy
	40	148	9.1	1023	13	ADR41584	ADR41584 Lipid acy
	41	147.5	9.0	12152	5	AA508699	AA508699 M. carbon
	42	147.5	9.0	37116	10	ABZ66810	ABZ66810 Orthomyx
	43	147	9.0	73882	13	ABD08026	ABD08026 Pseudomon
	44	146.5	9.0	2376	11	ABD08026	ABD08026 Pseudomon
	45	146	8.9	2120	10	ADG91050	ADG91050 Hepatic s

ALIGNMENTS

RESULT 1	AAAF6411	standard; DNA; 969 BP.
ID	AAAF6411	
XX	AAAF6411;	
AC	AAAF6411;	
XX		
DT	22-JUN-2001	(first entry)
XX		
DE	4-amino-4-deoxytrephenic acid dehydrogenase coding sequence.	
XX		
KW	Metabolite; benzene; chorismic acid; p-aminophenylpyruvic acid;	
KW	4-amino-4-deoxytrephenic acid dehydrogenase; enzyme; pAPC; ds.	
XX		
OS	Streptomyces venezuelae.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1..969
FT		/tag= a
FT		/product= "4-amino-4-deoxytrephenic acid dehydrogenase"
XX		
PN	MO200123542-A1.	
XX		
PD	05-APR-2001.	
XX		
PF	29-SEP-2000; 2000WO-JP006783.	
XX		
PR	29-SEP-1999; 99JP-00276314.	
XX		
PA	(MEIJ) MEIJI SEIKA KAISHA LTD.	

XX Yanai K, Okakura K, Yasuda S, Watanabe M, Miyamoto K, Midoh N;
 PI Murakami T;
 XX
 XX WPI; 2001-290517/30.
 DR P-PSDB; AAB82073.
 PT Transformant producing secondary metabolite modified with functional
 PT group e.g. benzene with nitrogen-containing substituent at para-position,
 PT PF1022, with ease at low cost, for application in pharmaceuticals and
 PT agrochemicals.
 XX
 PS Claim 15; Page 67-70; 83pp; Japanese.
 XX
 XX The present invention relates to a transformant having been modified so
 CC as to produce a secondary metabolite. The secondary metabolite has a
 CC benzene ring skeleton free from substitution at the para-position by a
 CC nitrogen-containing functional group, thereby enabling the production of
 CC a secondary metabolite with a benzene ring skeleton substituted at the
 CC para-position by a nitrogen-containing group. The transformant organism
 CC of the present invention has been produced by transferring a gene
 CC participating in the biosynthesis pathway from chorismic acid into p-
 CC aminophenylpyruvic acid. The present sequence is the coding sequence for
 CC 4-amino-4-deoxyphenolic acid dehydrogenase (pdc), from Streptomyces
 CC venezuelae. pdc participates in the biosynthesis pathway from chorismic
 CC acid into p-aminophenylpyruvic acid, and so the pdc gene can be used to
 CC produce the transformant of the present invention. The transformant can
 CC be used to produce metabolites for application in pharmaceuticals,
 CC veterinary drugs and agrochemicals
 XX
 SQ Sequence 969 BP; 106 A; 397 C; 358 G; 108 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 8.02e-88 Length: 969
 Score: 1635.00 Matches: 322
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 4 Gaps: 0

US-10-089-514-6 (1-322) x AAF6411 (1-969)

QY 1 MetSerGlyPheProArgSerValValValGlySerGlyAlaValGlyMetPhe 20
 DB 1 ATGAGCGGCTTCCCGCAGCGTCGTCGCGGAGCGGGCGGCGGCATGTC 60
 QY 21 AlaGlyLeuLeuArgGluAlaGlySerArgThrLeuValValAspLeuValProPro 40
 DB 61 GCCGGGCTGCTGGGAGGGGCGGACCGGACCGCTGCTGCACTCGTACCGGCGG 120
 QY 41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAla 60
 DB 121 GGAACGCGGAGCGCTGCTGCGGGGAGCTACCGCGCGGCGGCGGCACTCGGGCC 180
 QY 61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuTyr 80
 DB 181 GCCCTCGGGGACCGGAGCTCGCTGCTGCGCGGTACAGGACCTGGCGCTCAAGGCC 240
 QY 81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal 100
 DB 241 GTGGCGCGGTGACCGGCTCATGCGGCGGGGCGGCTGCTCGCCGACACCTGTCGTC 300
 QY 101 ArgThrGlyMetAlaIaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeu 120
 DB 301 CGGACGGGCGATGCGCGGAGCTCGGCGGCCACCGCGCTCCAGCACGCGGGCTTC 360
 QY 121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValThr 140
 DB 361 AACCCGATGTTCCGCCCGCGCGGCAATGACCGGCGGACCGCTGCGCGCGGTGTCACC 420
 QY 141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyValGlyPro 160
 DB 421 AGGAGCGGCGCGGCTGCAAGCGGCTGCTGCGGCTGCTGCAAGGCGGCGGCGGCGGCC 480

QY 161 ValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
 DB 481 GTACGGCTCAGCGCGGAGGACGACGCGGACGAGCGGCGGCCACCGGCGCTGACGCAC 540
 QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaGluGlyValAspValArgAlaLeu 200
 DB 541 GCCGTGCTCTCTCTCTCGGGCTCGCCCTCGCGCGCTGCGGCTGCGACCTCGGGCTTG 600
 QY 201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGly 220
 DB 601 GCGGCGAGGACCGCGCGGCCCAAGGTCGTCTGCGCTCTGCGCGCTGCTGCTGCGG 660
 QY 221 GlySerProGluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArg 240
 DB 661 GCGAGCGCGGAGGTGACGGGAGCATCCAGCGGTCCAAACCCCGGCGGCGTCCGCGCGC 720
 QY 241 ArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260
 DB 721 CGGGCGCTCGCGAGGCGCTGCTGCTGCGCGGCTGCTGCGGCGGACGACCGGACCT 780
 QY 261 AlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly 280
 DB 781 GCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
 QY 281 AsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGlyProGluLeuAlaGly 300
 DB 841 AACCTCGAGCGGCTCTTGGGGAGACTCGCGCGGTCAATGAGACCGGAGCTCGCGGGCG 900
 QY 301 GlnAspHisCysGlnGluLeuPheArgThrLeuHisAlaArgThrAspAspGluGly 320
 DB 901 CAGGACCACTGCGCAGAGAGCTGTTCCGACCTTCACCGGACCGGACGAGGCGAGAG 960
 QY 321 AspArg 322
 DB 961 GACCGA 966

RESULT 2
 AAL50182
 ID AAL50182 standard; DNA; 969 BP.
 XX
 AC AAL50182;
 XX
 DT 28-JAN-2003 (first entry)
 XX
 DE S venezuelae PF1022 substance gene #3.
 XX
 KW Substance PF1022; biosynthesis; transgenic; chorismic acid; anthelmintic;
 KW phenylalanine; p-aminophenylpyruvic acid; pharmaceutical;
 XX veterinary drug; gene; ds.
 XX
 OS Streptomyces venezuelae.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..969
 FT /*tag= a
 FT /product= "PF1022 substance"
 XX
 XX
 PN MO20027244-A1.
 XX
 PD 03-OCT-2002.
 XX
 PF 22-MAR-2002; 2002WO-JP002782.
 XX
 PR 22-MAR-2001; 2001JP-00082227.
 XX
 XX (MEIU) MEIU SEIKA KAISHA LTD.
 XX
 PA Yanai K, Sumida N, Watanabe M, Moriya T, Murakami T;
 PI WPI; 2003-018934/01.
 DR P-PSDB; AAO19565.
 XX

CC unnatural amino acid is biosynthesised by the cell. The present sequence
 CC represents a plasmid containing the individual genes papABC that encode
 CC the enzymes used to carry out the conversion of chorismate to the
 CC unnatural amino acid p-aminophenylalanine (pAF)

XX Sequence 3305 BP; 445 A; 1327 C; 1121 G; 412 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	1,49e-84	Length:	3305
Score:	1588.00	Matches:	315
Percent Similarity:	96.65%	Conservative:	2
Best Local Similarity:	96.04%	Mismatches:	5
Query Match:	97.13%	Indels:	6
DB:	8	Gaps:	1

US-10-089-514-6 (1-322) x ABZ69799 (1-3305)

```

QY      1 MetSerGlyPheProAArgSerValValGlyGlySerGlyAlaValGlyGlyMetPhe 20
DB      2007 ATGAGCGGCTTCCCGGAGCGCTGCTGCGGCGGAGCGGCGGCGGCGCATGTTTC 2066

QY      21 AlaGlyLeuLeuArgGluAlaGlySerArgThrLeuValAlaAspLeuValProProPro 40
DB      2067 GCCGGGCTGCTGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2126

QY      41 GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAla 60
DB      2127 GGAACGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2186

QY      61 AlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValGAla 80
DB      2187 GCCCTCCGGGACCGGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2246

QY      81 ValAlaProValThrArgLeuMetArgProGlyValAlaLeuLeuAlaAspThrLeuSerVal 100
DB      2247 GTGGCGGCTGTAACCGGCTCATGCAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2306

QY      101 ArgThrGlyMetAlaAlaGlyLeuAlaValHisAlaAlaProGlyValGlnHisValGlyLeu 120
DB      2307 CGGACGGGCGATGGCGGCGGAGCTCGGGGCCACCGCCCGGCGGCTCCAGCAAGCGGCGTC 2366

QY      121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValValThr 140
DB      2367 AACCCGATGTTCCCGCCCGCGGCGGCGGATGACCGGCGGCGGCGGCGGCGGCGGCTGTC 2426

QY      141 ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyArgProPro 160
DB      2427 AGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2486

QY      161 ValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
DB      2487 GTACGGCTCAACCGGCGGAGGACGACCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 2546

QY      181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValAlaGAlaLeu 200
DB      2547 GCCGTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2606

QY      201 AlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly 220
DB      2607 GCGGCGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2666

QY      221 GlySerProGluValArgGlyAspPheGlnArgSerSerAsnProArgAlaAlaSerAlaArg 240
DB      2667 GCGAGCCCGGAGGTGATCGGCGGACATCCAGCGGTCCACCCCGGCGGCGGCTCGGCGGCG 2726

QY      241 ArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAsp----- 256
DB      2727 CGGGCGCTCGCCAGGCGCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2786

QY      257 -----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274
DB      2787 GCCGAGGACCGGAGCGGCGGCGGAGACCCGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 2846
  
```

QY 275 GlyCysAspGlyAlaGlyAsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGly 294

DB 2847 GGATGCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2906

QY 295 ProGluLeuAlaAlaGlyValThrPheHisCysGlnGluLeuPheArgThrLeuHisArgThr 314

DB 2907 CCGAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2966

QY 315 AspAspGluGlyGlyLeuAspArg 322

DB 2967 GACGACGAAGCGGAGAGGACCGA 2990

RESULT 4

ABZ69798
 ID ABZ69798 standard; DNA; 12391 BP.

AC ABZ69798;

DT 08-APR-2003 (first entry)

DE Plasmid plasc-papabc.

KW DHFR, translation; orthogonal tRNA; O-tRNA; O-RS; CAT; TyrRS;

KW orthogonal aminoacyl tRNA synthetase; unnatural amino acid;

KW chloramphenicol acetyltransferase; dihydrofolate reductase; ds; plasmid;

XX biosynthesis; p-aminophenylalanine; pAF.

XX Synthetic.

XX WO200285923-A2.

XX 19-APR-2002; 2002WO-US012465.

XX 19-APR-2001; 2001US-0285030P.

XX 06-FEB-2002; 2002US-035514P.

XX (SCRI) SCRIPPS RES INST.

PI Schultz P, Wang L, Anderson JC, Chin JWK, Liu DR, Magliery TJ;

PI Meggers EL, Nehl RA, Pasternak M, Santoro SW, Zhang Z;

XX WPI; 2003-120430/11.

XX Composition useful for producing protein comprising unnatural amino acid,

PT has translation system comprising orthogonal tRNA and orthogonal

XX aminoacyl tRNA synthetase.

XX Example 4; Page 124-127; 1889p; English.

CC The invention relates to a novel composition comprising a translation

CC system comprising an orthogonal tRNA (O-tRNA) and an orthogonal aminoacyl

CC tRNA synthetase (O-RS), where the O-RS preferentially aminoacylates the O

CC tRNA with at least one unnatural amino acid in the translation system

CC and the O-tRNA recognises at least one selector codon. A composition of

CC the invention is useful for producing at least one protein comprising at

CC least one unnatural amino acid. The protein is the Asp127G mutant of

CC chloramphenicol acetyltransferase (CAT), the Tyr163Tg mutant of mouse

CC dihydrofolate reductase (DHFR), or a Tyr163Tg mutant of mouse

CC acid is provided exogenously. The translation system is a cell and the

CC unnatural amino acid is biosynthesised by the cell. The present sequence

CC represents a plasmid for use in the biosynthesis of p-aminophenylalanine

XX (pAF) in vivo

SQ Sequence 12391 BP; 2830 A; 3588 C; 3346 G; 2627 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	5,16e-84	Length:	12391
Score:	1588.00	Matches:	315
Percent Similarity:	96.65%	Conservative:	2
Best Local Similarity:	96.04%	Mismatches:	5

Query Match: 97.13% Indels: 6
 DB: 8 Gaps: 1
 US-10-089-514-6 (1-322) x ABZ69798 (1-12391)

QY 1 MetSerGlyPheProArgSerValValaIGlySerGlyAlaValaIGlyMetPhe 20
 DB 2361 ATGAGTGGCTTCCCGGAGCGTCTGTCGGCGGAGCGGAGCGGTCGTC 2420

QY 21 AlaGlyLeuLeuArgGlyAlaGlySerArgThrLeuValValaPheLeuValProProPro 40
 DB 2421 GCGGGGCTGCTGGCGGAGCGGCGAGCGCAGCCCTCTCTCCAGCCTCGTACCGCGCGCG 2480

QY 41 GlyArgProAspAlaCysLeuValaGlyAspValThrAlaProGlyProGlyLeuAlaAla 60
 DB 2481 GAGAGCGCGGAGCCCTGCTGTGTGGAGCTGACCGCGCGGCGCGGCTGCGGCG 2540

QY 61 AlaLeuArgAspAlaAspLeuValaLeuLeuAlaValaIsgIuAspValaAlaLeuVala 80
 DB 2541 GCGCTCCGGGAGCGGAGCCTCTCTCTGCTGCGGTCACAGAGACGTGGCCCTCAAGGCC 2600

QY 81 ValaIaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal 100
 DB 2601 GTGGCGCCGCTGAGCCGCGCTCATGCGAGCGGCGCGCTGCTCGCCAGCACCTGTCGCTC 2660

QY 101 ArgThrGlyMetAlaIaGlyLeuAlaAlaIsgIuAlaProGlyValaGlnIsgValaGlyLeu 120
 DB 2661 CGGACGGCGATGGCGCGGAGCTGCGCGGCGCGCGCGCGCGCGCGCTGCGCGCTC 2720

QY 121 AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValaAlaValaValaThr 140
 DB 2721 AACCCGATGTTCCCGCCCGCGCGCGATGACCGCGCGCGCGCGCGCGCTGCTGCTCACC 2780

QY 141 ArgAspGlyProGlyValaThrAlaLeuLeuArgLeuValaGlyGlyGlyArgPro 160
 DB 2781 AGGAGCGGCGCGCGCTGAGCGCGCTGCTGCGCTGCTGAGGAGCGCGCGCGCGCGCG 2840

QY 161 ValArgLeuThrAlaGlyGlnIsgAspArgThrThrAlaAlaThrGlnAlaLeuThrHis 180
 DB 2841 GTACGGCTCACCGCGGAGGAGCGAGCGCGGAGCGCGCGCGCGCGCGCTGAGCGCAC 2900

QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValaArgAlaLeu 200
 DB 2901 GCGGTATCTCTCTTCCGGGCTCGCGCTCGCGCGCGCTCGAGCTCCGGGCGCTG 2960

QY 201 AlaAlaThrAlaProProProIsgIuValaLeuLeuAlaLeuAlaArgValaGly 220
 DB 2961 GCGGCGAGCGGAGCGCGCGCGCGCGAGTGTCTGCGCTGCGCGCGCTGCTGCTGCG 3020

QY 221 GlySerProGlyValaThrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArg 240
 DB 3021 GCGAGCGCGGAGGTGAGCGGAGCATTCACCGGTCCAAACCCCGGCGCGGCTCGCGCG 3080

QY 241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValaGlyAsp 256
 DB 3081 CGGGCGCTCGCGGAGCGCTGCGCTTTCGCGCGCTGATCGGCGAGCGCGGAGCGCG 3140

QY 257 -----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274
 DB 3141 GCCGAGAGACCGGAGCGCGCGCGAGCGCGCGAGCGCGAGCAACCCCGGCGCGTCCCGGG 3200

QY 275 GlyCysAspGlyValaGlyAsnLeuAspGlyValaPheGlyGlyLeuArgArgLeuMetGly 294
 DB 3201 GAGTGGAGCGCGCGCGGAGCCTCGAGCGGCTTCGAGGAGACTCCGCGCGCTCATGGGA 3260

QY 295 ProGlyLeuAlaAlaGlyGlnAspHisCysGlnGlyLeuLeuPheArgThrLeuHisArgThr 314
 DB 3261 CCGGAGCTCGCGCGCGCGCGCGAGCACCACTGCGAGAGCTGTTCCGACCTTCACCGCAC 3320

QY 315 AspAspGlyGlyGlyValaArg 322
 DB 3321 GACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3344

RESULT 5
 AAT59269
 ID AAT59269 standard; cDNA; 888 BP.
 AC AAT59269;
 AC AAT59269;
 DT 02-APR-1997 (first entry)
 XX
 DE Streptomyces pristinaespiralis papC gene.
 XX
 KM Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
 KM DMPAP precursor; 4-dimethylamino-L-phenylalanine; papA; papM; papB;
 KM papC; isomerisation; aromatisation; N-methyltransferase; ds.
 OS Streptomyces pristinaespiralis.
 XX
 FH Key Location/Qualifiers
 FT mat_peptide 1..888
 FT /tag= a
 FT /product= "PapC"
 XX
 XX MO9601901-A1.
 XX
 XX 25-JAN-1996.
 XX
 PF 04-JUL-1995; 95WC-FR000889.
 XX
 PR 08-JUL-1994; 94FR-00008478.
 XX
 PA (RHON) RHONE POULENC ROBER SA.
 XX
 PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J,
 PI Barriere J, Debussche L, Famechon A, Paris J, Dutruc-Rosset G;
 DR MPI; 1996-097631/10.
 XX
 DR P-PSDB; AAW11582.
 XX
 PT New streptogramin B derivs. useful as antibiotics - produced by new
 PT mutants of Streptomyces having altered genes for streptogramin B
 PT biosynthesis.
 XX
 PS Claim 18; Page 105-106; 146pp; French.
 XX
 CC The papA and papM genes of *S. pristinaespiralis* are involved in the
 CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAP), a precursor for
 CC pristinamycin IA. The region between these two genes was sequenced and
 CC two open reading frames were identified. The first (papC) was on the
 CC complementary strand and decodes to an amino acid sequence with homology
 CC to the region of *E. coli* TyrA which has been implicated in aromatisation
 CC reactions. The papC gene product is likely to be involved in a similar
 CC phenylpyruvate during DMPAP synthesis. The second open reading frame
 CC (papB) could be decoded to give a product with homology to the region of
 CC TyrA which has chorismate mutase activity. The papB gene product is
 CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
 CC deoxy 4-aminoprephenate acid. Disruption of the papB and papC genes can
 CC be used to produce strains of *S. pristinaespiralis* which are unable to
 CC produce the antibiotic pristinamycin I but which may be able to produce
 CC new, modified forms of it
 XX
 SO Sequence 888 BP; 73 A; 254 C; 428 G; 133 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1.41e-30 Length: 888
 Score: 661.50 Matches: 155
 Percent Similarity: 60.38 Conservative: 34
 Best Local Similarity: 49.52 Mismatches: 101
 Query Match: 40.46 Indels: 23
 DB: 2 Gaps: 4

US-10-089-514-6 (1-322) x AAT59269 (1-888)

QY 2 SerGlyPheProArgSerValValaIGlySerGlyAlaValaIGlyMetPheAla 21

```

DB      13 TCAGTGTCTGGGCGTGTGTGTGTGGCGGGCGGCTGGCGCCGATGTTCAAC 72
QY      22 GtyleuLeuArgGluAgluArgThrLeuValValAspLeuValProProProGly 41
DB      73 CACTGCGTGTGGGCGTGGCGGCTGACCTGGCTGAC--GTGGCGGGGCGCGG 129
QY      42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuA 59
DB      130 GCGGCGGACCGGGGTGGGGGTGGCGGGATGTGGCGGCGCGGGCGGAGGGGCTC 189
QY      60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuLys 79
DB      190 GCGGCGCTGGCGGCGGCGGACGTGTGTCTGGCGGCTGGCGGCGGCGGTGGCGAG 249
QY      80 AlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer 99
DB      250 GCGGTGGAGGTGCTGGCGGGGGTGAATGGCGCCGCTGGCTGGCGGACACTTGTG 309
QY      100 ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly 119
DB      310 GTCAAGAGCGCGGATTCGCGGCGGCTGGCGAGCGGCGCGGGCTGCAGAGCGGTGGG 369
QY      120 LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaVal 139
DB      370 CTGAACCCGATGTTCCGCCCTGCTGGGTCTTCAAGGGGCGGCGGTGGCGGTGTG 429
QY      140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyValArg 159
DB      430 GTCAACGACGCGGCGCGGTGGCGGCGGCTGGTGAACCTGGCGGCGGGGGCGCGG 489
QY      160 ProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
DB      490 GTGTGGAGATGCGGCGGCGGCGGACAGACGACGACCGCGGCGAGCGCGGCGGACG 549
QY      180 HisAlaValaLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
DB      550 CATGCCGCGGTGCTGGCTTCCGGCGGCTGGCGGTGACGTTCCGTGGACGTGGCGGCGG 609
QY      200 LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
DB      610 CTGCGGACAGTCCCGCGCGGACGTGGCGGATGTGGCGTGGCGCGCGGATCGCC 669
QY      220 GlyGlySerProGluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla 239
DB      670 GCGGGGACCGCGGAGGTATTTGACATCAGCGCGGCAACCCCGCGCGCGCGCGGCGG 729
QY      240 ArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259
DB      730 CGGACGCGCTGGCGGCGGCGCTG-----GTGGCGCTGGGGCGAGCGCTCGAG 777
QY      260 ArgAlaAspAlaArgProGlyArgAlaAspAlaProGlyHisArgProGlyCysAspGlyAla 279
DB      778 AGGAGCGGACGAGAGACTTGGCC----- 801
QY      280 GlyAsnLeuAspGlyValPheGlyGluLeuArgArgLeuMetGlyProGluLeuAlaAla 299
DB      802 -----GCCCTGTTCCGCGAATCCGCGGTGTGTGGGCGAGACGCGCGCGAG 849
QY      300 GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis 312
DB      850 CTGGAACGCGTGTGGCGGCGGATGTTCAACGCGCCTGCAC 888

```

RESULT 6
AAT59268/c
ID AAT59268 standard; cDNA; 2888 BP.

AC AAT59268;
XX
DT 02-APR-1997 (first entry)
XX
DE Streptomyces pristinaespiralis papA and papM intergenic region.
XX

KW Streptogramin B; antibiotic; biosynthesis; pristinamycin; virginiamycin;
KM DMPAPA precursor; 4-dimethylamino-L-phenylalanine; PAPa; papM; papB;
KW papC; isomerisation; aromatisation; N-methyltransferase; ds.
OS Streptomyces pristinaespiralis.
XX
FH Key Location/Qualifiers
FT CDS 1..687
FT /tag= a
FT /product= "PAPa"
FT /note= "C-terminal coding region only, i.e. a partial
FT complement (949..1836)
FT CDS
FT /tag= b
FT /product= "papC"
FT 1873..2262
FT /tag= c
FT /product= "PAPa"
FT 2259..2888
FT /tag= d
FT /product= "papM"
FT /note= "N-terminal coding region only, i.e. a partial
FT open reading frame"

W09601901-A1.
25-JAN-1996.
04-JUL-1995; 95WO-FR000889.
08-JUL-1994; 94FR-00008478.
(RHON) RHONE POULENC RORER SA.

PI Blanc V, Thibaut D, Bamas-Jacques N, Blanche F, Crouzet J;
PI Barrière J, Debussche L, Famechon A, Paris J, Dutric-Rosset G;
DR WPI; 1996-097631/10.
P-PSDB; AAW11582, AAW11583.

PT New streptogramin B derivs. useful as antibiotics - produced by new
PT mutants of Streptomyces having altered genes for streptogramin B
PT biosynthesis.

XX Example 1; Page 102-104; 146gp; French.

CC The PAPa and papM genes of *S. pristinaespiralis* are involved in the
CC biosynthesis of 4-dimethylamino-L-phenylalanine (DMPAPA), a precursor for
CC pristinamycin IA. The region between these two genes was sequenced and
CC two open reading frames were identified. The first (papC) was on the
CC complementary strand and decodes to an amino acid sequence with homology
CC to the region of *E. coli* TyrA which has been implicated in aromatisation
CC reactions. The PAPa gene product is likely to be involved in a similar
CC aromatisation of 4-deoxy 4-amino prephenate to give 4-amino
CC phenylpyruvate during DMPAPA synthesis. The second open reading frame
CC (papB) could be decoded to give a product with homology to the region of
CC TyrA which has chorismate mutase activity. The PAPa gene product is
CC likely to catalyse the isomerisation of 4-deoxy 4-amino chorismate to 4-
CC deoxy 4-aminoprephenate. Disruption of the papB and papC genes can be
CC used to produce strains of *S. pristinaespiralis* which are unable to
CC produce the antibiotic pristinamycin I but which may be able to produce
CC new, modified forms of it

SQ Sequence 2888 BP; 390 A; 1319 C; 897 G; 282 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.:	4.27e-30	Length:	2888
Score:	661.50	Matches:	155
Percent Similarity:	60.38%	Conservative:	34
Best Local Similarity:	49.52%	Mismatches:	101
Query Match:	40.46%	Indels:	23
DB:	2	Gaps:	4

[illegible]

WP	ACF67367_12	1200001	1110000
WP	ACF67367_13	1300001	1410000
WP	ACF67367_14	1400001	1510000
WP	ACF67367_15	1500001	1610000
WP	ACF67367_16	1600001	1710000
WP	ACF67367_17	1700001	1810000
WP	ACF67367_18	1800001	1910000
WP	ACF67367_19	1900001	2010000
WP	ACF67367_20	2000001	2110000
WP	ACF67367_21	2100001	2210000
WP	ACF67367_22	2200001	2310000
WP	ACF67367_23	2300001	2410000
WP	ACF67367_24	2400001	2510000
WP	ACF67367_25	2500001	2610000
WP	ACF67367_26	2600001	2710000
WP	ACF67367_27	2700001	2810000
WP	ACF67367_28	2800001	2910000
WP	ACF67367_29	2900001	3010000
WP	ACF67367_30	3000001	3110000
WP	ACF67367_31	3100001	3210000
WP	ACF67367_32	3200001	3310000
WP	ACF67367_33	3300001	3410000
WP	ACF67367_34	3400001	3510000
WP	ACF67367_35	3500001	3610000
WP	ACF67367_36	3600001	3710000
WP	ACF67367_37	3700001	3810000
WP	ACF67367_38	3800001	3910000
WP	ACF67367_39	3900001	4010000
WP	ACF67367_40	4000001	4110000
WP	ACF67367_41	4100001	4210000
WP	ACF67367_42	4200001	4310000
WP	ACF67367_43	4300001	4410000
WP	ACF67367_44	4400001	4510000
WP	ACF67367_45	4500001	4610000
WP	ACF67367_46	4600001	4710000
WP	ACF67367_47	4700001	4810000
WP	ACF67367_48	4800001	4910000
WP	ACF67367_49	4900001	5010000
WP	ACF67367_50	5000001	5110000
WP	ACF67367_51	5100001	5210000
WP	ACF67367_52	5200001	5310000
WP	ACF67367_53	5300001	5410000
WP	ACF67367_54	5400001	5510000
WP	ACF67367_55	5500001	5610000
WP	ACF67367_56	5600001	5648894

Alignment Scores:

Pred. No.:

Score:

Percent Similarity:

Best local Similarity:

Query Match:

3.67e-12

382.00

47.94%

32.06%

23.36%

DB:

10

Gaps:

5

US-10-089-514-6 (1-322) x ACF67367_49 (1-110000)

OY

8

ValValValGlyGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla

27

Db

76463

GTCATCTCTTGGTGGGCAAGCGCGCATATGGTCTTATTTAGAGCGCCCTTTTACTCAATAT

76522

OY

28

GlySerArgThrLeuValAlaAspLeuValProProProGlyArg-----

42

Db

76533

GGTCATTTGTGATATCACTGTTCACAAACGACCGAGCTGCTTCATTAATCAT

76582

OY

43

ProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAlaLeu

62

Db

76583

TCTCATTAAGTCAGATT---GATATCTTAAACCCGCGTCGATGTGAGCGGTTTGG

76639

OY

63

ArgAspAlaAspLeuValLeuLeuAlaValAlaHisGlyIAspValAlaLeuYsAlaValAla

82

Db

76640

AGTGGGCGAAAGGCGGTCGCTTTCGCTTAACCGAAGCCGTCCTTAACAGCGTTGGCCT

76699

OY

83

ProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr

102

WP ADM27081_07 700001 810000
 WP ADM27081_08 800001 910000
 WP ADM27081_09 900001 1010000
 WP ADM27081_10 1000001 1110000
 WP ADM27081_11 1100001 1210000
 WP ADM27081_12 1200001 1310000
 WP ADM27081_13 1300001 1410000
 WP ADM27081_14 1400001 1510000
 WP ADM27081_15 1500001 1610000
 WP ADM27081_16 1600001 1694968

Alignment Scores:

Pred. No.: 1.09e-08 Length: 110000
 Score: 323.00 Matches: 93
 Percent Similarity: 52.55% Conservative: 41
 Best Local Similarity: 36.47% Mismatches: 104
 Query Match: 19.76% Indels: 18
 DB: Gaps: 8

US-10-089-514-6 (1-322) x ADM27081_07 (1-110000)

QY 6 ArgSerValValValGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuArg 25
 DB 59054 AGGATCGCGACTCTCGCGGTACCGGCATGAGCGCTTATCGCCCGGAGCTCCGT 59113
 QY 26 GluAlaGly-SerArg-----ThrLeuValAlaAspLeuValProProGlyArgPr 43
 DB 59114 GACGACGGTCATAGAGTACTGACTGTTCTAACCTCAGCCGCCGAGACCGCTCGCC 59173
 QY 43 oAspAlaCyLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAlaLeuArg 63
 DB 59174 CGCGACCTC-----GACCTCAGAGCT--GCTCCGACCAAGCTGATCCGCCAA 59220
 QY 63 gAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAla 83
 DB 59221 AGATGCCGACGTTGTGTTGTTCCGTCCTCCATTCGCTACGAAAGCTCATCCGTGA 59280
 QY 83 oValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrG 103
 DB 59281 GGTTCGACCAACAGCTTCCTGAGGTTCCCTCTCAGCAGTACCTCCCTCAAGGTTGC 59340
 QY 103 yMetAlaAlaGluLeuAlaAlaHisAlaPro--GlyValGlnHisValGlyLeuAsp 122
 DB 59341 TCCCGTTCGCGTATGCTCGAGCAGCTCCGAGATGTGTACGTCTCGAAGCTCAGCC 59400
 QY 122 oMetPheAlaProAlaAla--GlyMetThrGlyArgProValAlaAlaValAlaThr-- 140
 DB 59401 CTTCTTCGCGCCACGCGTCCCGAGTCTCCGCGGTACAGCCGTATCTCAAGCCACGGA 59460
 QY 141 -ArgAspGlyProGlyValThrAlaLeuValGlyLeuValGlyGlyValArgPr 160
 DB 59461 GCGTACTGTCCTCGAGCTCGCCGCGTCTCTTATCTGACGAGAAAGCCCTCAGT 59520
 QY 160 oValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrH 180
 DB 59521 CGTTGAGACCAACCCGAGAAACAGATCGCACTATGCGCGTTCAGTGCTCAACCA 59580
 QY 180 sAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArg-----LeuGlyValAs 196
 DB 59581 TGTCTGTCCTTCGTCGACGCGCCGCAATAGGCGCTTCTCCCAAGTTCGAGCTCGA 59640
 QY 196 pValArgAlaLeuAlaAlaThrAlaProProHisGlnValLeuLeuAlaLeuLeuAl 216
 DB 59641 CATCGAAGAGTCCGACG-----CCCGTATCGCTACTATGACGCTCGTGG 59691
 QY 216 aArgValLeuGlyGlySerProGluValGlyArgPheGlnArgSerAspProArgAl 236
 DB 59692 TCGTATCGCCGACGATCAGCTCTTATCGCGAGATCAAGCTTCAACCCCTAAG 59751
 QY 236 aAlaSerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPhe 250
 DB 59752 TGACGAAGCCCGTGAAGAGCTCTCGCGCCCTCAAGCGCTTC 59794

RESULT 11
 AAD50753
 AAD50753 standard; DNA; 1122 BP.

AC AAD50753;

DT 02-APR-2003 (first entry)

DE Escherichia coli bifunctional prephenate dehydrogenase DNA.

KW Tocopherol biosynthesis; prephenate dehydrogenase; chorismate mutase;
 KW tyra; transgenic; transgenic plant; homogenetic acid; plastoquinone;
 KW enzyme; gene; ds.

FT Escherichia coli.

FT Key Location/Qualifiers

FT CDS 1..1122
 /tag=a
 /product= "Bifunctional prephenate dehydrogenase"

PN MO200289561-A1.

PD 14-NOV-2002.

PF 03-MAY-2002; 2002WC-US013898.

PR 09-MAY-2001; 2001US-0289527P.

PA (MONS) MONSANTO TECHNOLOGY LLC.

PI Valentin HE, Miley TA;

PI MPI; 2003-058889/05.

DR P-PSDB; AAB32998.

XX Novel nucleic acid molecule encoding enzymes involved in tocopherol
 PT biosynthesis useful for enhancing or altering tocopherol production in
 PT plants such as canola, maize, soybean, crame, mustard, castor bean,
 PT peanut, sesame, or cottonseed.

PS Claim 5; Page 201-202; 206pp; English.

XX The present invention relates to novel nucleic acid molecules encoding
 CC enzymes involved in tocopherol biosynthesis. Polynucleotides of the
 CC invention comprise a promoter region functioning in a plant cell to cause
 CC production of an mRNA molecule linked to a heterologous nucleic acid
 CC molecule encoding an enzyme with prephenate dehydrogenase (tyra) or
 CC chorismate mutase activities or a tyra from Erwinia herbicola or
 CC Escherichia coli bifunctional prephenate dehydrogenase. Sequences of the
 CC invention are useful for producing a plant such as canola, maize,
 CC Arabidopsis, Brassica campestris, B.mapus, soybean, crame, mustard,
 CC castor bean, peanut, sesame, cottonseed, linseed, safflower, oil palm,
 CC flax and sunflower having increased tocopherol levels, especially alpha-
 CC or gamma-tocopherol or tocotrienol relative to a plant with a similar
 CC genetic background but lacking the exogenous nucleic acid molecule. They
 CC are useful for reducing tocopherol levels in a plant. They are also
 CC useful for producing plant expressing homogenetic acid and
 CC plastoquinones. The present sequence is Escherichia coli bifunctional
 CC prephenate dehydrogenase DNA

XX Sequence 1122 BP; 248 A; 242 C; 339 G; 293 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.66e-05 Length: 1122
 Score: 237.00 Matches: 78
 Percent Similarity: 45.39% Conservative: 50
 Best Local Similarity: 27.66% Mismatches: 110
 Query Match: 14.50% Indels: 44
 DB: Gaps: 8

US-10-089-514-6 (1-322) x AAD50753 (1-1122)

AAV21209_07
WP

8 Val1Val1Val1Gly1SerCylVal1Val1Gly1MetPheAlaGlyLeuThrArgLysIleA 27
|||
304 GTTATCTGTGGCGCGCGCGCATGAGACCTCTGTGCAGAAAGTCGTGACACTCCG 363
|||
28 GlySerArgThrLeuValValAspLeuValProProGlyArgProAspAlaCysLeu 47
|||
48 Val1GlyAspVal1Thr1AlaProGlyProGluLeuAla1AlaLeuArgAspAlaAspLeu 67
|||
406 GCGCGCATATTGTGCC-----GATGCGCGGAATG 435
|||
68 Val1LeuLeuAlaValHisGluAspVal1AlaLeuLysAlaVal1AlaProVal1ThraGlu 87
|||
436 GTGATTGTTAAGTGTCGCAATCCACGTACTGAGCAAGTTAATGCAATTCGCCCTTAA 495
|||
8 MetArgProGly1AlaLeuLeuAlaAspThrLeuSerValArgThrGly--MetAlaAla 106
:::
436 CCGAAA---GATTGTAATCTGGTGTATCTGGCATCTGATAAATGACCATTTACGGCC 552
|||
107 GluLeuAlaAlaHisAlaProGlyVal1GlnHisVal1GlyLeuAsnProMetPheAlaPro 126
|||||
553 ATGCTGGCGCGCACGATGCCCGTA-----CTGGGTTACACCANATGTTCCGTCG 606
|||
127 Ala1AlaGlyMetThrGlyArgProVal1Ala1AlaVal1ThraArgAspGlyProGlyVal 146
:::
607 GACACCGGTAGCGCTGGCAAGCAAGTT-----GTGCTGTGTGTATGACGCTAAACCG 660
|||
147 Thrala-----LeuLeuArgLeuVal1GluGly1Gly1ArgProVal1ArgLeu 163
|||
661 GAAGCATACCAATGCTTTTGGAGCAAATTCAGCTGGGGCGCTCGTTGCATCTGTAAT 720
|||
164 Thr1AlaGluGlnHisAspArgThrThra1Ala1AlaThrGlnAlaLeuThrHisAlaValLeu 183
:::
721 AGCGCGCTGACACGATCAGAATATGCGTTATTACAGCACTGCCACTTGTCTACT 780
|||
184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyVal1AspVal1ArgAlaLeuAla1Thr 203
:::
781 TTGTGCTTAGCGGCTGACCTGACAGAAAGAAATGTTCACTTGACGCAACTTCTGGCGTC 840
|||
204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly1GlySerPro 223
:::
841 TCTTGCCGATTTTACCGCTTGAGCTGGGAGATGCTGGCGCACTGTTCCCTCAGATCCG 900
|||
224 GluVal1YrGlyAsp1IleGlnArgSerAsnProArgAla1AlaSer1Ala----- 239
:::
901 CAGCTTTAAGCCGACATTATTATGCTGACAGCGTAATCTGCGCTTAATCAAACGTTAC 960
|||
240 -----ArgArgAlaLeuAla 244
|||
961 TATAAGCTTTTCGCGCAGCGCATGAGTTGCTGAGACAGGCGCATTAAGAGGCTTAAAT 1020
|||
245 GluAlaLeuArgSerPheAlaLeuVal1GlyAspAspProAspArgAlaAspAlaPro 264
:::
1021 GACACATTTCCGCAAGGTGAGACACTGTTCCGCGATTACGACACAGCGTTTTCAGACTGAA 1080
|||
265 GlyArg 266
|||
Db 1081 AGCCGC 1086

RESULT 12
AAV21209_09/c
Continuation (10 of 17) of AAV21209 from base 900001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS AAV21209 Accession Aav21209
Fragment Name Begin End
WP AAV21209_00 1 110000
WP AAV21209_01 100001 210000
WP AAV21209_02 200001 310000
WP AAV21209_03 300001 410000
WP AAV21209_04 400001 510000
WP AAV21209_05 500001 610000
WP AAV21209_06 600001 710000
WP AAV21209_07 700001 810000

WP	AAV21209_08	800001	910000	1
WP	AAV21208_09	900001	1010000	1
WP	AAV21209_10	1000001	1110000	1
WP	AAV21209_11	1100001	1210000	1
WP	AAV21209_12	1200001	1310000	1
WP	AAV21209_13	1300001	1410000	1
WP	AAV21209_14	1400001	1510000	1
WP	AAV21209_15	1500001	1610000	1
WP	AAV21209_16	1600001	1664976	1

Alignment Scores:

Pred. No.:	0.00164	Length:	110000
Score:	235.00	Matches:	71
Percent Similarity:	42.76%	Conservative:	59
Best Local Similarity:	23.36%	Mismatches:	158
Query Match:	14.37%	Indels:	16
DB:	2	Gaps:	5

US-10-089-514-6 (1-322) x AAV21209_09 (1-110000)

QY	9	valValaGlyserGlyalValaGlyGlymetPhealGlyLeuLeuargGluGly	28
Db	30033	ATTATTGAGGAACTAGTGTATTAGGGAAGTGTGTCGAAGATTTTAAATAAAGG	29974
QY	29	SerargThrLeuValValaAspLeuValProProGlyArgProAspAlaCysLeuVal	48
Db	29973	TTTAACCTTATGTTATCTGCGAGAGATATTGAGAAAGAAATTTTGAGAAAGATT	29914
QY	49	GlyAspValThrIaProGlyProGlyLeuLeuAlaAlaLeuargAspIaAspLeuVal	68
Db	29913	GGG--GTGAATTACTTAACAACACATTGAAGTGT--AAAAGAGATATTGT	29860
QY	69	LeuLeuAlaValHisGlyAspValAlaLeuValaValaIaProValThrArgLeuMet	88
Db	29859	ATTGTACAGTTCACATTAATGTTTACGAAGGTTATATAAGAGTACTCTCATGTT	29800
QY	89	ArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaGlyLeu	108
Db	29799	AGGGAAGAGTGTATTATTAATGACATPAACCTCATTTAAAGAGATTCCTCAAAAGCTATG	29740
QY	109	AlaIaHisAla---ProGlyValGlnHisValGlyLeuAspProMetPheAlaProAla	127
Db	29739	GAAAGAGATGTTAAAGAGGAGTTACAGTTATTCCAACCCACCAATGTTGGCCCATCA	29680
QY	128	AlaGlyMetThrGlyArgProValAlaAlaValaValThr-----ArgAspGly	143
Db	29679	ACACCTTCTTTGTTTAACAACAAGTGTATCTTAAACCCCTTCGAAAGCATAAATACT	29620
QY	144	ProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyValArgProValArgLeu	163
Db	29619	GAGTGTGTTAAACAAGGTTTATTAATTTTAAAGAAAGAGGGGCTAAGGTTATATGTCATC	29560
QY	164	ThrIaGlyGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu	183
Db	29559	CCTTCACAAAAACATGTATGAATTAATGAGGATTTGTTCAAGGTTTGACTCACTCGCTTT	29500
QY	184	LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr	203
Db	29499	ATATCTTTAGAGCAACGTTAAAAAGAACTCAACGTGATTTATAAAGATCAAGAAAGTTT	29440
QY	204	AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerPro	223
Db	29439	GCCGCCCAATATATACGAGTGTGATGATTTGATCATTTGGGAGGTTTATGACAGAAATCC	29380
QY	224	GluValIaGlyAspIleGlnArgSerAspProArgAlaAlaSerAlaArgArgAlaLeu	243
Db	29379	TATTTATATCTGACATCCAAAGTTTATCCAAAGATTAAGAGATTCATGAACCTTT	29320
QY	244	AlaGlyIaLeuArgSerPheAlaAlaLeuValGlyAspProSerArgIaAspAla	263
Db	29319	ATTAATATAGGTATAGGAATACAGTGAATTTGTTAAAAATTAAGATTAAGAAAGTTTGT	29260
QY	264	ProGlyArgIaAspAlaProGlyHisProGlyGlyCysAspGlyAlaGly-----	280

```
Db      29259 AAGATAATGAAAGACCTGCAAGCATTGTTGTAAGTGAAGCAAGAGGGGCTTATTAT 29200
      281 -----AsnLeuAepGlyValPheGlyGluLeuArgArgLeuMetGly 294
Db      29199 TCAGATTAAGCCGCTGTTTGCAATTACATCTGAAATTGAAAGCTTAATAACTCATTTGGT 29140
      295 ProGluLeuA1a 298
      29139 AAAGATGTGGCT 29128

RESULT 13
ABN83448
ID      ABN83448 standard; DNA; 1122 BP.
XX
AC      ABN83448;
XX
DT      27-AUG-2002 (first entry)
XX
DE      Escherichia coli prephenate dehydrogenase coding sequence.
XX
KW      Arogenate dehydrogenase; enzyme; herbicide; tyrosine biosynthesis;
      prephenate dehydrogenase; gene; ds.
XX
OS      Escherichia coli.
FH
FH      Key      Location/Qualifiers
FT      CDS      1..1122
FT              /*tag= a
FT              /product= "Prephenate dehydrogenase"
XX
XX      WO200246441-A2.
XX      13-JUN-2002.
XX      PD
XX      05-DEC-2001; 2001WO-FR003832.
XX      PR
XX      05-DEC-2000; 2000FR-00015723.
XX      PA
XX      (AVET ) AVENTIS CROPS SCIENCE SA.
XX
XX      PI      Matringe M, Ripert P;
XX      DR      WPI; 2002-500449/53.
XX      DR      P-PSDB; AB83256.
XX
XX      PT      New nucleic acid encoding arogenate dehydrogenase, useful for identifying
XX      herbicides, also related protein and herbicide-tolerant plants.
XX
XX      PS      Disclosure; Page 58-60; 66pp; French.
XX
XX      CC      The present invention relates to arogenate dehydrogenase (ADH) enzymes
XX      (see ABN83439-ABN83443 and ABN83248-ABN83251). The ADH enzymes catalyse
XX      the last stage of the metabolic pathway of tyrosine biosynthesis and are
XX      potential targets of herbicides. The present sequence is the coding
XX      CC      sequence for prephenate dehydrogenase, which was used to illustrate the
XX      invention
XX
XX      SQ      Sequence 1122 BP; 243 A; 245 C; 342 G; 292 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:      2,866-05      Length:      1122
Score:      233.00      Matches:      77
Percent Similarity: 45.04%      Conservative: 50
Best Local Similarity: 27.30%      Mismatches: 111
Query Match: 14.25%      Indels: 44
DB:      6      Gaps: 8

US-10-089-514-6 (1-322) x ABN83448 (1-1122)
QY      8 ValValValGlyGlySerGlyValValGlyGlyMetPheAlaGlyLeuLeuArgGluA1a 27
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      304 GTTATCTGCGCGGTGCGGTGCGATGAGGACGCGCTGTTGAGAAAGATGCTGACCTCTCG 363
```

```
QY      28 GlySerArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      364 GGTATTCAGTGCAGTGCAGTTCGGAG-----CAACATACCTGGAGTTCGA 405
QY      48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      406 GCGGCTGATATGTTGCC-----GATGCCGGAATG 435
QY      68 ValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeu 87
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      436 GTGATTTGATGTTGTCCAATCCAGTTACTGAGCAAGTTATTGGCAATTCACGCTTTA 495
QY      88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      496 CCGAAA---GATTGATTTCTGTCGATCTGCGATCAGTGAATAATGGCGCATTCACGGCC 552
QY      107 GluLeuAlaAlaHisLeuAspGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      553 ATGCTGTGCGCGCATGATGTCCTCGTG-----CTGGGCGTACACCCGATGTTCCGTCG 606
QY      127 AlaAlaGlyMetThrGlyArgProValAlaAlaValValThrArgAspGlyProGlyVal 146
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      607 GACAGCGGTAGCCTGGCAAGCAAGTT-----GTGCTGTGATGAGCACTAAACCG 660
QY      147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      661 GAAGCATACCAATGTTTCTGGAGCAAAATTCAGCTGTGGCGGCTGCGTCATCTATT 720
QY      164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      721 ACGCGCGTGCAGACGATCAGATATGCGCTTATTACAGGACCTGGCCGCACTTGCTACT 780
QY      184 LeuSerPheGlyLeuAlaLeuValArgLeuGlyValAspValArgAlaLeuAlaIleTrp 203
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      781 TTTCCTTACGGCTGCACCTGGCAGAAAGAAATGTTACCTTGAGCACTTCTGGCGCTC 840
QY      204 AlaProProPheGlnValLeuLeuAlaLeuValaArgValLeuGlySerPro 223
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      841 TCTTGGCGATTTTACCGCTTACGCTTGAGCTGGGAGTGTGGCGGCTGCTCAGATCCG 900
QY      224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      901 CAGCTTATGCGCATCATCTATGTGTCAGACGCTAATCTGCGTTAATCAAAACGTTAC 960
QY      240 -----ArgArgAlaLeuAla 244
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      961 TATTAAGCTTTCGGCGAGCGATTAAGTTGCTGAGACAGGCGGCAATTAAGCGCTTATT 1020
QY      245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
      |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::
      1021 GACAGTTTCCGCAAGGTGAGACATGCTTGGCGCATTCAGCACACCGCTTTTCAGACTGA 1080
QY      265 GlyArg 266
      |||
      1081 ACGCCG 1086

RESULT 14
ABA9612
ID      ABA9612 standard; DNA; 1238 BP.
XX
XX      ABA9612;
XX
DT      17-MAY-2002 (first entry)
XX
DE      E. coli tyra DNA.
XX
XX      Shikimate pathway; chorismate mutase; tyra; prephenate dehydrogenase;
XX      vitamin E; vitamin K; ubiquinone; fat-soluble antioxidant; transgenic;
XX      feed; pharmaceutical; cosmetic; aromatic amino acid; salicylic acid;
XX      folic acid; phenylpropanoid; flavanoid; plant; abiotic stress resistance;
XX      tocopherol; tocotrienol; gene; ds.
```

```

OS Escherichia coli.
XX Key Location/Qualifiers
FH CDS 25..1146
FT /*tag= a
FT /product= "tyrA"
XX
XX MO200200901-A1.
XX
XX 03-JAN-2002.
XX
XX 28-JUN-2001; 2001MO-EP007391.
XX
XX 29-JUN-2000; 2000DE-01030647.
XX 21-DEC-2000; 2000DE-01064454.
XX
XX (SUNG-) SUNGENE GMBH & CO KGAA.
XX
XX Badur R, Geiger M, Kunze I, Sommer S;
XX WPI; 2002-164442/21.
XX P-PSDB; AAM49653.
XX
XX Preparing fine chemicals, particularly Vitamins E and K, useful as
XX antioxidants e.g. in foods or medicine, by growing organisms with altered
XX shikimate biosynthesis pathway.
XX
XX Claim 20; Page 55-57; 74pp; German.
XX
XX This invention describes a novel method for the preparation of fine
XX chemicals by culturing organisms in which the shikimate pathway has been
XX altered relative to the wild type. The method involves a construct
XX containing a plastid transit peptide, a chorismate mutase and/or
XX prephenate dehydrogenase, linked to at least one regulatory sequence for
XX transcription or translation in plants. The method is used to produce
XX chemicals, particularly Vitamins E or K and/or ubiquinone, useful e.g. as
XX fat-soluble antioxidants (in feeds, pharmaceuticals and cosmetics), but
XX also aromatic amino acids, salicylic or folic acid derivatives.
XX
XX Phenylpropanoids, flavanoids etc., especially in transgenic plants. Also
XX plants with increased Vitamin E contents have improved resistance to
XX abiotic stress, e.g. frost and drought. Transgenic plants with an altered
XX shikimate pathway are useful as foods, fodder and in preparation of
XX processed foodstuffs. Transgenic plants with modified shikimate pathways
XX have increased content of desired chemicals, particularly tocopherols
XX and/or tocotrienols. This sequence encodes the Escherichia coli tyrA
XX (chorismate mutase prephenate dehydrogenase) protein used in the method
XX of the invention
XX
XX Sequence 1238 BP; 269 A; 275 C; 370 G; 324 T; 0 U; 0 Other;
XX
XX
XX Alignment Scores:
XX Pred. No.: 3,14e-05 Length: 1238
XX Score: 233.00 Matches: 77
XX Percent Similarity: 45.04% Conservative: 50
XX Best Local Similarity: 27.30% Mismatches: 111
XX Query Match: 14.25% Indels: 44
XX DB: Gaps: 8
XX
XX US-10-089-514-6 (1-322) x ABA99612 (1-1238)
XX
XX 8 ValValValGlyGlySerGlyAlaValAlaGlyGlyMetPheAlaGlyLeuAlaGlyGluAla 27
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 328 GTTATCTCGCGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 387
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 28 GlySerArgThrLeuValValAlaLeuValProProProGlyArgProAspAlaCysLeu 47
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 388 GGTATCATCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 429
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaAlaLeuArgAspAlaAspLeu 67
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 430 GCGGCTGATATTTGTTGCC-----GATGCCGGAATG 459
XX |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 68 ValLeuLeuAlaValAlaHisGluAspValAlaLeuValAlaValAlaProValThrArgLeu 87

```

```

DB 460 GTGATTGTAGTGTGCGCAATCCAGCTTACTGAGCAAGTATTGGCAAAATTACCGCCTTTA 519
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 520 CCGAAA---GATTGTATCTGTGCGATCTGCGCATGTGCAAAAATGCGCATACGGCC 576
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 577 ATGCTGTGCGCGATATGATGTCGCGT-----CTGAGGCTACACCGAGTTGCGTCCG 630
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 631 GACAGCGGTAGCCTGCGCAAGCAAGTT-----GTGCTGCGTGTATGACATAAGCAACG 684
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 147 ThrAla-----LeuLeuArgLeuValGlyGlyGlyValArgProValArgLeu 163
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 685 GAAGCATACCAATGTTCTGCGCAAAATTCAGTCTGCGGCGCTGCGCATCGTATT 744
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrAlaAlaLeuThrHisAlaValLeu 183
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 745 AGCGCGGTGAGCAGCATCAGATATGCGCTTTATTCAGGCACTGCGCATTTGCTACT 804
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 805 TTTGCTTACGGGCTGCACTGCGCAAGAAATGTTACAGTTAGCAACTTTCGCGCTC 864
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 204 AlaProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGlyGlySerPro 223
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 865 TCTTCCCGCATTTACCGCCTTGAGCTTGCGGATGTCGCGGCACTTGTCTCAGATCCG 924
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 925 CAGCTTATGCCGACATCATATATGTGTGACAGCGTAATCTGCGTTAAACGATTAC 984
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 240 -----ArgArgAlaLeuAla 244
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 985 TATTAACGTTTCCGCGCAGCGATTTGCTGTCGAGCAGGCGCATTAAGCGGCTTTATT 1044
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1045 GACAGTTTCCGCAAGGTGAGCAGCTGTCGCGGATTTAGCAGCGTTTTCAGAGTAA 1104
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 265 GlyArg 266
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DB 1105 AGCGCG 1110
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
RESULT 15
AAL46032
ID AAL46032 standard; DNA; 1238 BP.
XX
XX AAL46032;
XX
XX 11-JUL-2002 (first entry)
XX
XX E coli TyrA gene.
XX
XX
XX Vitamin E; homogenisate; HG; homogenisate-1,2-dioxygenase; HGD;
XX maleylacetate-isomerase; MAI; fumarylacetate-hydrolase; FAH;
XX transgenic; nutrition; antioxidant; cardiovascular disease; cancer;
XX immune system; generalised age-related degeneration; animal feed;
XX meat quality; cosmetics; growth regulator herbicide; cardiac;
XX cytosolic; immunostimulant; enzyme; TyrA; chorismatmutase;
XX prephenate dehydrogenase; gene; db.
XX
XX
XX Escherichia coli.
XX
XX Key Location/Qualifiers
FH CDS 25..1146
FT /*tag= a
FT /product= "bifunctional chorismatmutase/
FT prephenate dehydrogenase"
XX

```

PN WO200231173-A2.
 XX
 PD 18-APR-2002.
 PF 18-SEP-2001; 2001WO-EP010779.
 XX 19-SEP-2000; 2000DE-01046462.
 PR
 XX (SUNG-) SUNGENE GMBH & CO KGAA.
 PA
 XX Geiger M, Ebneeth M, Kunze I;
 PI
 XX WPI; 2002-362682/39.
 DR P-PSDB; AA017421.
 XX
 PT Increasing Vitamin E production, useful in human or animal nutrition,
 PT comprises reducing activity of enzymes that metabolize homogenitase.
 PS
 XX Disclosure; Page 65-67; 98pp; German.
 CC The present invention relates to a method of producing Vitamin E,
 CC involving modulating the synthesis of vitamin E by reducing decomposition
 CC of homogenitase (HG) by reducing the activity of homogenitase-1,2-
 CC dioxygenase (HGD), maleylacetate-isomerase (MAI) and/or
 CC fumarylacetate-hydrolase (FAH). The method is used to produce
 CC transgenic organisms that are useful in animal and human nutrition and
 CC for the isolation of vitamin E. Vitamin E is a fat-soluble antioxidant
 CC with a protective effect against cardiovascular disease and cancer, it
 CC also stimulates the immune system and may prevent generalised age-related
 CC degeneration. When used in animal feeds, it improves quality and
 CC storability of meat and can also be used in cosmetics. Also antibodies
 CC raised against the 3 specified enzymes are useful in screening for
 CC specific inhibitors, potentially useful as growth regulators, e.g.
 CC herbicides. The present sequence is the E coli Tyra gene, which encodes
 CC the bifunctional chorismatase/prephenatdehydrogenase enzyme
 XX
 XX Sequence 1238 BP; 269 A; 275 C; 370 G; 324 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 3,14e-05 Length: 1238
 Score: 233.00 Matches: 77
 Percent Similarity: 45.04% Conservative: 50
 Best Local Similarity: 27.30% Mismatches: 111
 Query Match: 14.25% Indels: 44
 DB: 6 Gaps: 8

US-10-089-514-6 (1-322) x AA146032 (1-1238)

QY 8 ValValValGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuLeuArgGluAla 27
 DB 328 GTTATCGTCGCGCGGTGCGGTGAGGAGCGCTGTCGAGAAATGTCGACCTCTCG 387
 QY 28 GlySerArgThrLeuValValAlaPheLeuValProProProGlyArgProAlaCysLeu 47
 DB 388 GGTATTCAGGTGCGGATTCGAG-----CAACATGACCTGGATCGA 429
 QY 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
 DB 430 GCGGCTGATTTGTTGCC-----GATGCCGAATG 459
 QY 68 ValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeu 87
 DB 460 GTGATTGTAGTGTCCCAATCCAGCTTATGACCAAGTTATGGCAATTAACGCCCTTA 519
 QY 88 MetArgProGlyValAlaLeuLeuAlaAspThrLeuSerValArgThrGly--MetAlaAla 106
 DB 520 CCGAAA---GATTGTTTCGTGTCGATCTGCGCATCGTGAATAATGGCCATTACAGGCC 576
 QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
 DB 577 ATGCTGTGCGCGCATGATGTCGCGTG-----CTGGGCTACACCCGATGTTGCGTCCG 630
 QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaValThrArgAspGlyProGlyVal 146

DB 631 GACAGCGGTAGCCTGCGCAAGCAAGTT-----GTGCTCGTGTGATGACGTAAACCG 684
 QY ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeu 163
 DB 685 GAAGCATACCAATGTTTCTTGAGCAAAATTCAGTCTGGGCGCTGCGTCATCGTATT 744
 QY ThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
 DB 745 AGCGCGGTGAGACATCAGAAATATGCGCTTATTCAGGCACTGGCCACTTGTGCTACT 804
 QY LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
 DB 805 TTGCTTACGGGCTGCACCTGCGCAAAATGTTCACTTGACCAACTTTCGCGCTC 864
 QY AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerPro 223
 DB 865 TCTTCCCGCATTTTACCGCTTGAGCTGGCGATGTCGGCGACCTGTTGCTCAGATCCG 924
 QY GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla----- 239
 DB 925 CAGCTTATGCGACATCATTTATGTCGTGAGCGTAATCTGCGCTTAATCAACCTTAC 984
 QY 240 -----ArgArgAlaLeuAla 244
 DB 985 TATTAAGCTTTCGCGAGCGCATTTGAGTTGTCGAGCAGCGCATTAAGCAGCGTTTATT 1044
 QY 245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
 DB 1045 GACAGTTTCCCGAAGGTGAGCATGTTGCGGATTAACGACAGCGTTTTCAGAGTGAA 1104
 QY 265 GlyArg 266
 DB 1105 AGCCGC 1110

Search completed: October 5, 2005, 10:51:48
 Job time : 728.456 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 10:07:08 ; Search time 176.216 Seconds
(without alignments)
2989.973 Million cell updates/sec

Title: US-10-089-514-6
Perfect score: 1635
Sequence: 1 MSGFPRSVVVGSGAVGGMF.....HCOELPRTLHRTDDEKXDR 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 1202784 seqs, 818138359 residues

Total number of hits satisfying chosen parameters: 2405568

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=x1h
-Q=/cgn2_1/USPTO_spool_h/US10089514/runat_04102005_105746_8071/app_query.fasta_1.1621
-DB=Issued Patents NA -OPMT=faastap -SUFFIX=rnt1 -MINMATCH=0.1 -LOOPTCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blonum62 -TRANS=humano4.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTPMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514.@CGN_1.116 @runat_04102005_105746_8071 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARM TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:*
1: /cgn2_6/ptodata/1/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	661.5	40.5	888	US-08-765-907A-2	Sequence 2, Appl1
2	661.5	40.5	888	US-09-987-614A-2	Sequence 2, Appl1
3	661.5	40.5	2888	US-08-765-907A-1	Sequence 1, Appl1
4	661.5	40.5	2888	US-09-987-614A-1	Sequence 1, Appl1
5	235	14.4	1664976	US-08-916-421B-1	Sequence 1, Appl1
6	235	14.4	1664976	US-09-692-570-1	Sequence 1, Appl1
7	227	13.9	1131	US-09-489-039A-4572	Sequence 4572, Ap
8	216	13.2	1830121	US-09-557-884-1	Sequence 1, Appl1
9	216	13.2	1830121	US-09-643-990A-1	Sequence 1, Appl1
10	200	10.2	1125	US-09-543-681A-1138	Sequence 1138, Ap
11	166.5	10.2	1038	US-09-902-540-5637	Sequence 5637, Ap
12	166.5	10.2	49225	US-09-902-540-1269	Sequence 1269, Ap

C	13	163.5	10.0	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
	14	150.5	9.2	1362	4	US-09-724-797-31	Sequence 31, Appl1
C	15	149	9.1	2732	4	US-09-589-892B-3	Sequence 3, Appl1
	16	149	9.1	3810	4	US-09-902-540-3021	Sequence 3021, Ap
C	17	149	9.1	3812	4	US-09-902-540-641	Sequence 641, Ap
	18	147	9.0	1848	4	US-09-902-540-3350	Sequence 3320, Ap
C	19	147	9.0	19302	4	US-09-902-540-1155	Sequence 1155, Ap
	20	146.5	9.0	2376	4	US-09-902-540-1155	Sequence 6630, Ap
C	21	146	8.9	3263	4	US-09-252-991A-5310	Sequence 5310, Ap
	22	146	8.9	3262	4	US-09-252-991A-5256	Sequence 5256, Ap
C	23	146	8.9	4158	4	US-09-252-991A-5348	Sequence 5348, Ap
	24	146	8.9	4953	4	US-09-252-991A-5348	Sequence 5227, Ap
C	25	145	8.9	3600	1	US-08-537-002A-5	Sequence 5, Appl1
	26	145	8.9	3600	1	US-08-863-010-5	Sequence 5, Appl1
C	27	145	8.9	3600	3	US-09-024-429-5	Sequence 5, Appl1
	28	145	8.9	3957	4	US-10-237-551-193	Sequence 193, Ap
C	29	145	8.9	154746	4	US-09-827-688-8	Sequence 8, Appl1
	30	145	8.9	154746	4	US-09-827-688-8	Sequence 8, Appl1
C	31	144	8.8	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
	32	141.5	8.7	1527	4	US-09-248-796A-4088	Sequence 4088, Ap
C	33	140	8.6	21010	4	US-09-902-540-1188	Sequence 1188, Ap
	34	140	8.6	4403765	3	US-09-103-840A-2	Sequence 2, Appl1
C	35	140	8.6	4411529	3	US-09-103-840A-1	Sequence 1, Appl1
	36	139.5	8.5	1289	3	US-09-385-028-21	Sequence 21, Appl1
C	37	139.5	8.5	1289	4	US-09-726-614-21	Sequence 21, Appl1
	38	139.5	8.5	1289	4	US-09-385-028-21	Sequence 21, Appl1
C	39	139.5	8.5	11604	3	US-09-385-028-13	Sequence 13, Appl1
	40	139.5	8.5	11604	4	US-09-726-614-13	Sequence 13, Appl1
C	41	139.5	8.5	11604	4	US-09-385-040-13	Sequence 13, Appl1
	42	139.5	8.5	15079	3	US-09-385-028-1	Sequence 1, Appl1
C	43	139.5	8.5	15079	4	US-09-726-614-1	Sequence 1, Appl1
	44	139.5	8.5	15120	4	US-09-385-040-1	Sequence 1, Appl1
C	45	139	8.5	2175	4	US-09-252-991A-15144	Sequence 15144, A

ALIGNMENTS

RESULT 1
US-08-765-907A-2
; Sequence 2, Application US/08765907A
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BEMAS-JACQUES, Nathalie
; APPLICANT: BIANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUISSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; TITLE OF INVENTION: Streptogramins
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765, 907A
; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 888
; TYPE: DNA
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-2

Alignment Scores:

Pred. No.: 4,81e-51
Score: 661.50
Percent Similarity: 60.38%
Best Local Similarity: 49.52%
Query Match: 40.46%
DB: 3
Length: 888
Matches: 155
Conservative: 34
Mismatches: 101
Indels: 23
Gaps: 4

```

US-10-089-514-6 (1-322) x US-08-765-907A-2 (1-888)

Oy      2 SerGlyPheProAArgSerValValAGlyGlySerGlyAlaValAGlyMetPheAla 21
Db      13 TCGGTGTTCCGGGGGTTGTGTGTGTGGCGGGGCGGGCGGTGGCGCGGATTTTACG 72
Oy      22 GlyLeuMetArgGlnAGlySerArgThrLeuValAlaAspLeuValProProGly 41
Db      73 CACTGCGTGTGGCGTTCCGGGGGTCGGCGTGAACCTGGCTGAC--GTGGCGGGCGGCT 125
Oy      42 ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGlyLeuAla 59
Db      130 GCGGCGGACGGGGGCGGGGCGGTGGTGGCGGATGTGGCGGCGCGGCGCGGAGCGGTC 189
Oy      60 AlaAlaLeuArgAspAlaAspLeuValLeuLeuAlaValHisGlyAspValAlaLeuLys 79
Db      190 GCGGCGCTGTGGCGGGCGGCGGACGTGTGTGTGTGGCGGTGGCCGAGACCGGTGGCGGAG 249
Oy      80 AlaValAlaProValThrArgMetArgProGlyValAlaLeuLeuAlaAspThrLeuSer 99
Db      250 GCGGTGGAGGTGCTGGCGGGGGGAGTGGCGGACCGGGCGGTGTCTGGCGGACCTTGTGG 309
Oy      100 ValArgThrGlyMetAlaAGlyLeuAlaValHisAlaProGlyValAGlnHisValAGly 119
Db      310 GTCAAGAGCCGGATTCGCGGGCGGCGTGGCGTGAAGCGCGCGCGGGCTCGACAGCGGTGGG 366
Oy      120 LeuAsnProMetPheAlaProAlaAlaAGlyMetThrGlyArgProValAlaAlaValVal 139
Db      370 CTGAACCCGATGTTCGCGCCCTCGCGGTCTTCACAGGGCGGCGCGGTGGCGCGGTGGGTG 422
Oy      140 ThrArgAspGlyProGlyValThrAlaLeuLeuAlaGlyLeuValGlnGlyGlyArg 155
Db      430 GTCAACCAACGGGCGCGGTGTGGCGGGCTGTGTGTGTGTGGCGGGTGGGGGGCGCGG 489
Oy      160 ProValArgLeuThrAlaGlnGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
Db      490 GTGTGTGAAGATGCCGGCGCGCGCGGACACAGACGTGACCGCGCGGACAGCGGCGGACG 549
Oy      180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
Db      550 CATCCCGCGGTGTGGCTTCGGGGCTGGGCGTGGGTGAGCTGTGGTGAAGTGTGGGGGCGG 609
Oy      200 LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
Db      610 CTGGGGGACAGTCCCGCGCGCGCATCTGGCGATGTGGCGCTGTGGCCCGCATGCC 668
Oy      220 GlyGlySerProGlnValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAla 239
Db      670 GCGCGGACGCGCGGAGGTGTATTTTCACATCCAGCGCGGACCGCGCGGCGCGCGCGGCG 729
Oy      240 ArgArgAlaLeuAlaGlnAlaLeuAlaLeuArgSerPheAlaAlaLeuValAGlyAspAspProAsp 259
Db      730 CCGGAGGCGCTGGGCGCGGCGCTG-----GTGGCGGTGGGCGAGGCGGTGGAG 777
Oy      260 ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAla 279
Db      778 AGGGGCGACACAGAGAGAGCTTCCGCG----- 801
Oy      280 GlyAsnLeuAspGlyAlaPheGlyGlnLeuAlaArgArgLeuMetGlyProGlnLeuAlaAla 299
Db      802 -----GCCCTGTGTCCCGGAATCGCGGGGTGTGTGGGGGACAGCGGCGGAG 849
Oy      300 GlyGlnAspHisCysGlnGlnLeuPheArgThrLeuHis 312
Db      850 CTGGAACGCGTGTGGCGCGGATGTTCAACGCGGCTGTGCAC 888

RESULT 2
US-09-987-614A-2
; Sequence 2, Application US/09987614A
; Patent No. 6833382
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis

```

```

APPLICANT: BAMAS-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBOUSCHE, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
TITLE OF INVENTION: Multaaynthesis
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/09/987, 614A
CURRENT FILING DATE: 2001-11-15
PRIOR APPLICATION NUMBER: US/06/765, 907
PRIOR FILING DATE: 1997-03-20
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 888
TYPE: DNA
ORGANISM: Streptomyces pristinaespiralis
US-09-987-614A-2

```

Alignment Scores:	
Pred. No.:	4,81e-51
Score:	661.50
Percent Similarity:	60.38%
Best Local Similarity:	49.52%
Query Match:	40.46%
BB:	4
Length:	888
Matches:	155
Conservative:	34
Mismatches:	101
Indels:	23
Gaps:	4

US-10-089-514-6 (1-322) X US-09-987-614A-2 (1-888)

QY	2	SeqGLYPheARSerServalValValGlyGlySerGlyValValGlyGlyMetPhePhala	21
DB	13	TGGGTGTCGGCGCTTGTCGTGTGTGGTGGCGGGCCGGGTGGGTGGGCCCATGTTCAAC	72
QY	22	GlyLeuLeuArgGluAlaGlySerArgThrLeuValValAlaPheLeuValProProProGly	41
DB	73	CACGGCTGTGTCGTTCCGGCGGTGGCGGGACCTGGCTGGAC---GTGGCCGGGGCGGT	122
QY	42	ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGluLeuAla	59
DB	130	GGCGCGGAGCGGGGTGGGGGTGGCTGGCCGGTATGTGGCCGGCGGGCCGGAGCGGTC	188
QY	60	AlaAlaLeuArgAspAlaAspLeuValLeuLeuValValHisGluAspValAlaLeuLys	79
DB	190	GGCGGCGCTGCGCGCGCGGACCTGGTGTGGTCTGGCGGGTCCGAGCCGGTGGCTGGAG	248
QY	80	AlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSer	99
DB	250	GGCGTGGAGAGTCTGGCGGGGTGTATGGGCCCGGTGGCTCTCGCGACACTTGTTCG	308
QY	100	ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly	119
DB	310	GTCAGAGCGCGATGCGCCGGCGGCTGGCTGAGCGCGCGCGCGGAGCTGCAGCGGTGGG	368
QY	120	LeuAspProMetPheAlaProAlaAlaGlyMetThrGlyAspProValAlaAlaValAl	139
DB	370	CTGACCCCATGTTGCCCTTCCTGGTGGATCTTACAGGGGGCCGGGTGGCCGGTGGTG	428
QY	140	ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArg	158
DB	430	GTCACCGAGCGGGCCGGGTGTGGGGCCCTGTGTGAGAGCTGTGGCGGGTGGGGGGCCGG	488
QY	160	ProValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr	179
DB	490	GTTGTGTGAATGCCGGCGCGCGGACAGACAGCTGACCGCGCGGACAGCCGACG	548
QY	180	HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAlaAspValArgAla	199
DB	550	CATGGCGCGTCTGCTCTGGAGTGGGCGCTGGGTGAGCTGTGGGTGAGAGTGGGGGCG	609

```

Oy      200  LeuAlaAlaThAlaProProRoHiAGlnValIleuLeuAlaLeuAlaAGValLeu 219
           :::::|||||
Db      610  CTGCGGAGACAGTGTCCCGCGCGCGCATCTGGCGATGTGGCTGTGGCGCGGATCGCC 669
           :::::|||||
Oy      220  GluGlySerProGlnValTyrGlyAspIleGlnArgSerAspProArgAlaAlaSerAla 239
           :::::|||||
Db      670  GCGCGGAGACGCGGAGGTTATTTTCGACATCTCGACGCGCGCAACCGCGCGCGCGCGCGG 729
           :::::|||||
Oy      240  ArgArgAlaLeuAlaGlnAlaLeuAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp 259
           :::::|||||
Db      720  CGGAGAGCGCTGTGGCGCGCGCGCTG-----CTGCGCGCTGGGCGAGCGCGCTCGAG 777
           :::::|||||
Oy      260  ArgAlaAspAlaAspProGlyAlaArgAlaAspAlaProGlyHisPProGlyGlyCyAspArgAla 279
           :::::|||||
Db      778  AGGGGCGAGACGAGGACGTTCCG----- 801
           :::::|||||
Oy      280  GluAspLeuAspArgValPheGlyGlyValArgArgLeuMetGlyProGlyLeuAlaAla 299
           :::::|||||
Db      802  -----GCCCTGTTCGCGGAACTGCGGAGTGTCTGTGGCGAGACGCGCGCGGAG 849
           :::::|||||
Oy      300  GluGlnAspRhiLeCyGlnGlnLeuPheArgThLeuHis 312
           :::::|||||
Db      850  CTGGAAAGGCTGTGTGCGCGGAGATGTTCCAGCGCGCTGAC 888
           :::::|||||

```

RESULT 3

```

US-08-765-907A-1/C
: Sequence 1, Application US/08765907A
: Patent No. 6352639
:
: GENERAL INFORMATION:
: APPLICANT: BLANC, Veronique
: APPLICANT: THIBAUT, Denis
: APPLICANT: BAMA-JACOUES, Nathalie
: APPLICANT: BLANCHE, Francis
: APPLICANT: COUZET, Joel
: APPLICANT: BARRIERE, Jean-Claude
: APPLICANT: DEBUSCHÉ, Laurent
: APPLICANT: FAMECHON, Alain
: APPLICANT: PARIS, Jean-Marc
: APPLICANT: DUTRUC-ROSSET, Gilles
: TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
: TITLE OF INVENTION: Mutasynthesis
: FILE REFERENCE: Streptogramin genes
: CURRENT APPLICATION NUMBER: US/08/765,907A
: CURRENT FILING DATE: 1997-03-20
:
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 1
: LENGTH: 2888
:
: TYPE: DNA
: ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-1

```

Alignment Scores:

Pred. No.:	2,48e-50	length:	288
Score:	661.50	Matches:	155
Percent Similarity:	60.38%	Conservative:	34
Best Local Similarity:	49.52%	Mismatches:	101
Query Match:	40.46%	Indels:	23
DB:	3	Gaps:	4

US-10-089-514-6 (1-322) X US-08-765-907A-1 (1-2888)

QY	2	sergilypneproargserValValGlyglyserGlyAlaValGlyIglymetchPhala	21
Db	1824	TCGGATTCCGGCCGTTGTGTGTGTGTGGCGGGCGCGGTGGCCGATGTCAGC	1765
QY	22	GlyLeuLeuArgGluAlaGlyserArgThrLeuValValAspLeuValProProGly	41
Db	1764	CACTCGCTGGTGCCTTCGGGGGTGGCGGTGACCTGGCTGAC--GTGACCGGGCCCGT	1708
QY	42	ArgProAspAlaCysLeuVal-----GlyAspValThrAlaProGlyProGlyLeuAla	59
Db	1707	GCGCGAGACGGGGTCCGGGTGGTGGCCCGATGATGTGGCGCGCGCGCCGAGAGCGGTC	1648

QY	60	AlaIalaleuArgAspAlaAspLeuValleuLeuAlaValHisGluAspValAlaLeuLys	79
Db	1647	GGGGGGCTGGCGGGCGGAGAGTGTGTGTGTGTGGCGGATCCGGAGCCGGGTGGCTGGGAG	1588
QY	80	AlaValAlaProValThrArgLeuMetCArgProGlyAlaLeuLeuAlaAspThrLeuSer	99
Db	1587	GGGGTGGAGGTGTGGGGGGGGGTATATCGGGCCGGGTCCGGGTCTCGGGACACTTGTTCG	1528
QY	100	ValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGly	119
Db	1527	GTCAAGAGCCGGATCGCCGGCGGCTCGTAGCGCGCGCGGGGCTGGCAGGCCGTGGGG	1468
QY	120	LeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValVal	139
Db	1467	CTGAACCCGAGTGTGGCCCCCTCGCTGGGTCTTACAGAGGGCGGCGGTGGCGGCTGTG	1408
QY	140	ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArg	159
Db	1407	GTACACGACGGGGCCGGGTGTGGCGGCTGTGGAGCTGGTGCGCGGGTGGGGGGCCGG	1348
QY	160	ProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr	179
Db	1347	GTGGTGGAGATGCGCGGGCGGCGGCGGACAGACAGCTGACCCCGGCGGACAGGCGCACG	1288
QY	180	HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla	199
Db	1287	CATCCCGGGGTGGCTTGGCTGGGCTGGGCTGGGTAGACTGTGGTGGAGCTGGGGGCG	1228
QY	200	LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu	219
Db	1227	CTGGGGGACAGTGGCCCGCGCGGCACCTGGCGCATGTGGGGGTGCTGGCCGGATGGCC	1168
QY	220	GlyGlySerProGluValIYrGlyAspIleGlnArgSerAsnProAlaAlaAspValA	239
Db	1167	GGCGGGACGCGCGGAGGTGTATTTTCGACATCCAGGCGCCCAACCCGCGCGCGCGCG	1108
QY	240	ArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAsp	259
Db	1107	CGGACGGGGCTGGGGCGCGGCTCT-----GTGGGGCTGGGGGACAGCGCTGGAG	1060
QY	260	ArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyAla	279
Db	1059	AGGGGCGACGAGAGAGCTTGGC-----	1036
QY	280	GlyAsnLeuAspGlyValPheGlyGlnLeuArgArgLeuMetCglyProGluLeuAlaAla	299
Db	1035	-----GCCCTGTTCTGCCGAACCTGGCGCGTGTGGCTGGGCGGAGCACGGCGCGAG	988
QY	300	GlyGlnAspHisCysGlnGluLeuPheArgThrLeuHis	312
Db	987	CTGGACACCGCTGTGCGCGGAGATTTCACCGCCCTTGAC	949

RESULT 4

US-09-987-614A-1/C
Sequence 1, Application US/09987614A
Patent No. 6833382
GENERAL INFORMATION:
APPLICANT: BLANC, Veronique
APPLICANT: THIBAUT, Denis
APPLICANT: BAMA-JACQUES, Nathalie
APPLICANT: BLANCHE, Francis
APPLICANT: COUZET, Joel
APPLICANT: BARRIERE, Jean-Claude
APPLICANT: DEBUSCHÉ, Laurent
APPLICANT: FAMECHON, Alain
APPLICANT: PARIS, Jean-Marc
APPLICANT: DUTRUC-ROSSET, Gilles
TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
TITLE OF INVENTION: Metasynthesis
FILE REFERENCE: Streptogramin genes
CURRENT APPLICATION NUMBER: US/09/987,614A
CURRENT FILING DATE: 2001-11-15


```
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995) .. (191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980) .. (231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187) .. (234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220) .. (234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814) .. (234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398) .. (309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418) .. (309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837) .. (312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993) .. (312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226) .. (319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167) .. (559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241) .. (559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992) .. (600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708) .. (622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081) .. (657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203) .. (657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435) .. (674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442) .. (682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652) .. (713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684) .. (741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455) .. (779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676) .. (779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539) .. (855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619) .. (871619)
OTHER INFORMATION: n equals a, t, c, or g
```

```
NAME/KEY: misc_feature
LOCATION: (1084830) .. (1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846) .. (1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881) .. (1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881) .. (1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988) .. (1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224) .. (1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473) .. (1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491) .. (1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091) .. (1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020) .. (1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1602912) .. (1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1603734) .. (1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1637998) .. (1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1664854) .. (1664854)
OTHER INFORMATION: n equals a, t, c, or g
```

US-08-916-421B-1

```
Alignment Scores:
Pred. No.: 9,71e-08 Length: 1664976
Score: 235.00 Matches: 71
Percent Similarity: 42.76 Conservative: 59
Best Local Similarity: 23.36 Mismatches: 158
Query Match: 14.37 Indels: 16
DB: 4 Gaps: 5
```

US-10-089-514-6 (1-322) x US-08-916-421B-1 (1-1664976)

```
QY 9 ValValIGlVlySerClYlAlaValIGlYgMctPheAlGlyLeuLeuArgGluIaGly 28
DB 930033 ATTATTGAGAGAACTGATGCTTTAGGGAAGTGGTTGCAAGATTATTAATAAGG 929974
QY 29 SerArgThrLeuValAlaLeuValAlaLeuValAlaLeuValAlaLeuValAlaLeuVal 48
DB 929973 TTTAACGTTATAGTTACTGCGAGAGATATTGAGAAAGAAAAGTTAGAAAGCTTA 929914
QY 49 GlYAsPValThrAlaPrGlyPrGluLeuAlaAlaAlaLeuArgAspAlaAspLeuVal 68
DB 929913 GGG---GTTGAATTTACTAACCAACAACATTTGAAGCTGCT---AAAAAGGAGATATTGTT 929860
QY 69 LeuLeuAlaValAlaHisGluAspValAlaLeuValAlaLeuValAlaLeuValAlaLeuVal 88
DB 929859 ATTGTAGCAGTTCCAAATTAATGTTACAGAAAGGTTTAAAGAGTACTCTCATGTT 929800
QY 89 ArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaIaGluLeu 108
DB 929799 AGGGAAGATGTTTATTATATGACATTAACCTCAATTAAAGATTCCTCAAGAGCTATG 929740
```


	OTHER INFORMATION:	n	equals	a,	t,	c,	or	g
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(309418) .. (309418)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(312637) .. (312637)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(312993) .. (312993)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(319226) .. (319226)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(559167) .. (559167)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(559241) .. (559241)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(600992) .. (600992)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(622708) .. (622708)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(657081) .. (657081)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(657203) .. (657203)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(674435) .. (674435)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(682442) .. (682442)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(713652) .. (713652)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(741684) .. (741684)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(779455) .. (779455)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(779676) .. (779676)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(855539) .. (855539)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	
FEATURE:								
NAME/KEY:	misc_feature							
LOCATION:	(871619) .. (871619)							
OTHER INFORMATION:	n	equals	a,	t,	c,	or	g	

```

FEATURE:
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1313224)..(1313224)
OTHER INFORMATION: n equals a, t, c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1349473)..(1349473)
OTHER INFORMATION: t, c, or g

```

Alignment Scores:

Alignment Score:	
Pred. No.:	9,71e-08
Score:	235.00
Percent Similarity:	42.76%
Best Local Similarity:	23.36%
Query Match:	14.37%
DB:	4
	5
Length:	1664976
Matches:	71
Conservative:	59
Mismatches:	158
Indels:	16
Gaps:	5

US-10-089-514-6 (1-322) x US-09-692-570-1 (1-1664976)

9 ValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28

Db 930033 ATTATTGAGGAACTGATGCTTTAGGGAAGTGGTTTGCAGATATTTAAAAAATAAAGCG 929974

29 SerArgThrLeuValValAspLeuValProProGlyArgProAspAlaCysLeuVal 488

Db 929973 TTAAAGTAAAGTACTGGAGAGATATTGAGAAAGGAAAAATGTTGAGAAAGAGTTA 929914

49 G I Y A S P V A I I n t r a l a p r o g l y p r o g l u l e u a l a l a l e u a r g a s p a l a a s p l e u a l 68

DB 929913 GGG---GTCAGATTACCAACAACATGAGCTGCT---AAAAAGGAGATATGT 92

09 LEUUEWATAVAI NIBOI WASPVAI AIALEUUYA NIAVAI AIAPI OVAI I MIAIGREUMEC 08
::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: :::: ::::

[illegible]

x_1

[illegible]

二
 三
 四
 五
 六
 七
 八
 九
 十
 十一
 十二
 十三
 十四
 十五
 十六
 十七
 十八
 十九
 二十
 二十一
 二十二
 二十三
 二十四
 二十五
 二十六
 二十七
 二十八
 二十九
 三十
 三十一
 三十二
 三十三
 三十四
 三十五
 三十六
 三十七
 三十八
 三十九
 四十
 四十一
 四十二
 四十三
 四十四
 四十五
 四十六
 四十七
 四十八
 四十九
 五十
 五十一
 五十二
 五十三
 五十四
 五十五
 五十六
 五十七
 五十八
 五十九
 六十
 六十一
 六十二
 六十三
 六十四
 六十五
 六十六
 六十七
 六十八
 六十九
 七十
 七十一
 七十二
 七十三
 七十四
 七十五
 七十六
 七十七
 七十八
 七十九
 八十
 八十一
 八十二
 八十三
 八十四
 八十五
 八十六
 八十七
 八十八
 八十九
 九十
 九十一
 九十二
 九十三
 九十四
 九十五
 九十六
 九十七
 九十八
 九十九
 一百

[illegible][illegible]

On 144 Prog]vYa]Thra]a]u]u]A]ra]u]u]a]G]u]G]u]G]u]A]ra]P]ro]v]a]A]ra]u] 163

Db 929619 GAGTGGTTTAAACAAGGTTTATATTTTAAAGAAGAGGGCTTAAGGTTATAGTCATTC 929560

164 ThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 18

 艮 山

```
Db 929559 CCTCCAGAAAAACATGATAGAAATTATGCGGATTTGTCAGGGTTGACTCATTCGCTTT 929500
Qy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaIthr 203
Db 929499 ATATCTTAGAGCAACGTTTAAAGAACTCAACGTTGATATTAAGAGTCAAGAACTTT 929440
Qy 204 AlaProProHISglnValLeuLeuAlaLeuAlaIthrValLeuGlySerPro 223
Db 929439 GCCTCCCAATATACGAGTTGATGATTTCTATCATTTGGAGGATTTATAGGACGAATCCC 929380
Qy 224 GluValTyrGlyAspIleGlnArgSerAspArgAlaIleSerAlaArgAlaLeu 243
Db 929379 TATTTATATCTGACATCCAAATGTTTAAATCAAGGTAAGAGATTCTATGAACCTTT 929320
Qy 244 AlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAla 263
Db 929319 ATAAATCATGTATAGCAAAATCAGTGAATTTCTTAAAAATAAAGATAGAGAGTTTGT 929260
Qy 264 ProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyValagly----- 280
Db 929259 AAGATATAGAAAGACGTCGAAAGCATTTTGTAGTACGCAAGAGAGGGCTTATAT 929200
Qy 281 -----AenLeuAspGlyValPheGlyGluLeuArgAlaGluMetGly 294
Db 929199 TCAGATAGAGCCGCTGTTGCAATTAAACATCTGAAATGGAAGCTAAATAACTCATTTGT 929140
Qy 295 ProGluLeuAla 298
Db 929139 AAAGATGTGCT 929128

RESULT 7
US-09-489-039A-4572
; Sequence 4572, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489, 039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4572
; LENGTH: 1131
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4572

Alignment Scores:
Pred. No.: 2,03e-11 Length: 1131
Score: 227.00 Matches: 83
Percent Similarity: 45.80% Conservative: 48
Best Local Similarity: 29.02% Mismatches: 107
Query Match: 13.82% Indels: 49
DB: 4 Gaps: 9

US-10-089-514-6 (1-322) x US-09-489-039A-4572 (1-1131)
Qy 8 ValValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
Db 313 GTATGCTGTCGGCGCGGCGGAGATGCGCGCTGTTGAGAAATGTTAAAGCTGTTC 372
Qy 28 GlySerArgThrLeuValAlaAspLeuValProProGlyArgProAspAlaCysLeu 47
Db 373 GGCCTACCAAGGTGCGTATCTGAA-----AAAAACGATTGGCCCGCG 414
Qy 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaIleLeuArgAspAlaAspLeu 67
Db 415 GCGCGGATATCGTCC-----GATGCCGCGCATG 444
Qy 68 ValLeuLeuAlaValHisGluAspValAlaLeuLysAlaValAlaProValThrArgLeu 87
```

```
Db 445 GTTATGTCAGCGTACCGATTCAACACACCGTGGAGACATGCGAGCGCTGCCGCCCTT 504
Qy 88 MetArgProGlyValAlaLeuAlaAspThrLeuSerValArgThr---GlyMetAlaIle 106
Db 505 ---CCGCGGATTCGATTCGTGTTGACCTTGCTCGGTAAAGCGAGCCGCCGCGGG 561
Qy 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
Db 562 ATGCTGGCGGCGGATCATGAGGCGCGGT-----CTTGGCTGCACCCGATGTTCCGTCG 615
Qy 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Db 616 GACAGCGGAGCGCTGGCGAAGCAGGTG-----GTGCTATATGTATGATGCGCGCACGCG 669
Qy 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Db 670 GAGGCTTATCAGTGTCTCTCGAACAATCCAGTTTGGGGCGCGGCTGCACCGCATC 729
Qy 164 ThrAlaGluGlyHisAspArgThrThrAlaAlaIthrGlnAlaLeuThrHisAlaValLeu 183
Db 730 AGGCGCGTTGAGCAGCATCAGAACATGCGCTTTATTCAGGCGCTGCGCATTTGCGCAC 789
Qy 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaIthr 203
Db 790 TTCGCTATGCGCTGCACCTGCGGAGAAATGTACGTCTTGAGCAACTGCTGCGGCTC 849
Qy 204 AlaProProHISglnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerPro 223
Db 850 TCGTCAACCATTTTACCGACTGAGCTGGCGAGTGTAGGCGCTGTTGCCCATGATCCG 909
Qy 224 GluValTyrGlyAspIleGlnArgSerAspProArgAlaAlaSerAlaArgAlaLeu 243
Db 910 CAGCTTACGCGATTTTATTCATGTCGGA-AAACAACCGGCGCTATTAAAGCGTA 968
Qy 244 AlaGluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAla 263
Db 969 CTACACGCTTTCCG-----GAGCAGCGCGGATAGCA 1019
Qy 264 ProGlyArgAlaAspAlaProGlyHisProGlyGlyCysAspGlyValaglyAsnLeuAsp 283
Db 984 -----CGAAGCATGCGCTGCT-----GAGCAGCGCGGATAGCA 1019
Qy 284 GlyValPheGlyGluLeu 289
Db 1020 GCGCTTATCGACAGCTT 1037

RESULT 8
US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: MD
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3 1/2 inch diskette
; COMPUTER: Dell Pentium
; OPERATING SYSTEM: MS DOS v6.22
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/557,884
; FILING DATE: 25-Apr-2000
; CLASSIFICATION: <Unknown>
```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-557-884-1

Alignment Scores:

Pred. No.:	5.9e-06	Length:	1830121
Score:	216.00	Matches:	77
Percent Similarity:	45.26%	Conservative:	47
Best Local Similarity:	28.10%	Mismatches:	118
Query Match:	13.21%	Indels:	32
DB:	4	Gaps:	8

US-10-089-514-6 (1-322) x US-09-557-884-1 (1-1830121)

QY 2 SerGlyPheProArgSerValValGlyGlySerGlyValAlaValGlyMetPheAla 21
|||
|||
Db 1370739 TCTGATATTCACAAATTTGTTATTTGGCGGTTATGCGGCTTAATTAATTTGGC 1370798
|||
|||
QY 22 GlyLeuLeuArgGlyLalaGlySerArgThrLeuValAlaSerLeuValProProGly 41
|||
|||
Db 1370799 CGTTATTACGTGCTGCTGCTCAATTCTTAATTTTACAT----- 1370840
|||
|||
QY 42 ArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuAlaAla 61
|||
|||
Db 1370841 CGGAAAGATTGGCGGCTGCTGAAAGATT----- 1370870
|||
|||
QY 62 LeuArgAspAlaAspLeuValLeuLeuAlaAlaHisGlyAspValAlaLeuValAla 81
|||
|||
Db 1370871 TTAGCGAATGCTGATGTCGATGCTTCCGCTTATTAATCTCACTTGAAGAACTT 1370930
|||
|||
QY 82 AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal--- 100
|||
|||
Db 1370931 GAACGCTTAAACCTTATTTAACGAAACATGCTACTTCAAGATTTAACCTCTTAAG 1370990
|||
|||
QY 101 ArgThrGlyMetAlaAlaGlyLeuAlaAlaHisAlaAlaProGlyValGlnHisValGlyLeu 120
|||
|||
Db 1370991 CGGAAACGCTGCGGAAATGCTTGAAGTTCAATGCTGCTCTT-----TTAGGTTTA 1371044
|||
|||
QY 121 AspProMetPhe---AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
|||
|||
Db 1371045 CATCCATGTTTGTGTCAGATATTCAGAGTATGCGAAACAAAGTGTGTGCTTGTGAT 1371104
|||
|||
QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyArg 159
|||
|||
Db 1371105 GACGCTTTCTGTAACGTTATGATGTTACTTGACGAATTCAAATTTGGGCTGCAAAA 1371164
|||
|||
QY 160 ProValArgLeuThrLalaGlyGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
|||
|||
Db 1371165 ATTTATCAAAACCAATGCGACACAGATCAATATATGACTTAATACAGCTTGGCGC 1371224
|||
|||
QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
|||
|||
Db 1371225 CATTTTTCAGCTTTGCGATGTTTACACTTTCCAAAGCCCATTAATCTCGCTAAT 1371284
|||
|||
QY 200 LeuAlaAlaThrAlaProProGlnHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
|||
|||
Db 1371285 TTATTGGCACTTTCTTCCCTATTATTCGGTTAGACTTGCATGATAGTCCGTTATATT 1371344

QY 220 GlyGlySerProGlyValThrGlyAspIle-----GlnArgSerAspProArgAlaAla 237
|||
|||Db 1371345 GCCCAAGATGACAGGCTTTACGAGATATTATTATGATTAAGTCAGAAATTTAGCGGCTA 1371404
|||
|||QY 238 -----SerAlaArgArgAlaLeuAlaGlyLalaLeuArgSerPheAlaAlaLeuValGly 255
|||
|||Db 1371405 ATTGAACCTTAACAACTTACGACGAACTTAACTTCTTT----- 1371449
|||
|||QY 256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
|||
|||Db 1371450 -----GAAATATATGATGCTCAAGGTTTATTATGATGCT 1371482
|||
|||

RESULT 9

US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289GENERAL INFORMATION:
APPLICANT: Robert D. FleischmannMark D. Adams
Owen White
Hamilton O. Smith

J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of the Haemophilus influenzae Rd Genome, Fragments Thereof, and Uses Thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MDCOUNTRY: USA
ZIP: 20850COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch disketteCOMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/643,990A

FILING DATE: 23-Aug-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429

FILING DATE: 1995-06-07

APPLICATION NUMBER: 08/426,787

FILING DATE: 1995-04-21

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover

REGISTRATION NUMBER: 40,302

REFERENCE/DOCKET NUMBER: PB186P1C1

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790

TELEFAX: 310-309-8439

INFORMATION FOR SEQ ID NO: 1:
LENGTH: 1830121 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Alignment Scores:

Pred. No.:	5.9e-06	Length:	1830121
Score:	216.00	Matches:	77
Percent Similarity:	45.26%	Conservative:	47
Best Local Similarity:	28.10%	Mismatches:	118
Query Match:	13.21%	Indels:	32
DB:	4	Gaps:	8

US-10-089-514-6 (1-322) x US-09-643-990A-1 (1-1830121)

```
QY      2 SerGlyPheProArgSerValValGlyGlySerGlyValAlaValGlyMetPheAla 21
      1370739 TCTGATATATTCACAAATTTGTTATGTCGGCGTTATGCGGCTTATTTGCGC 1370798
      22 GlyLeuLeuArgGluAlaGlySerArgThrLeuValValAspLeuValProProProGly 41
      1370799 CGTTATTTAGCTCATCTGCGTATCCAAATTTCTATTATTACAT----- 1370840
      QY      42 ArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaAla 61
      1370841 CGGAGATGTTGGCGCGTGAAGATTT----- 1370870
      QY      62 LeuArgAspAlaAspLeuValLeuLeuAlaValAlaHisGluAspValAlaLeuValAlaVal 81
      1370871 TTAGCAATGCTGATGTCGATGCTTTCGCTCCATTATTAATCTCACTTGAACAAAT 1370930
      QY      82 AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal--- 100
      1370931 GAACGCTTAAACCTTATTTAACGGAACAAATGCTACTGCAATTTAACCTCTGTTAAG 1370990
      QY      101 ArgThrGlyMetAlaAlaGlyLeuAlaAlaHisAlaProGlyValAlaGlnHisValGlyLeu 120
      1370991 CGTGAACCGCTACGCAAAAGCTTTGAAGTTCAATCGTGTCTGTT-----TTAGGTTTA 1371044
      QY      121 AsnProMetPhe--AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
      1371045 CATCCATGTTGTTGTCAGATATTTGCAAGTATGCAAAACAAAGTGTGTCGCTTGAT 1371104
      QY      140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArg 159
      1371105 GGACGTTTTCCTGAACGTTATGTAATGATTACTTGACCAAAATTTGGGGTGCAGAA 1371164
      QY      160 ProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
      1371165 ATTATCAACCAATGCGCAAGAACACGATCAATATATGACTATATACAGCTTGGCGC 1371224
      QY      180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
      1371225 CATTTTTCGACTTTCGATGAGTGTTAACCTTTCCAAACAGCCCATTAATCTCGCTAAT 1371284
      QY      200 LeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
      1371285 TATATGCACTTTCCTCCCTATTTATCGGTATGACCTGCGATGATGCTCGTTATTT 1371344
      QY      220 GlyGlySerProGluValTyrGlyAspIle-----GlnArgSerAsnProArgAlaAla 237
      1371345 GCGCAAGATGCAAGCTTTACGCAATATTATATGATATAGTCACAGAAATTTAGCGGTA 1371404
      QY      238 -----SerAlaArgArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGly 255
      1371405 ATTGAACCGCTAAACAACTTACGACGAAAGCTTAACTTTCTT----- 1371449
      QY      256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
      1371450 -----GAAATAATGATGCTCAAGGTTTATTATGATGCT 1371482
      RESULT 10
      US-09-543-681A-1138
      ; Sequence 1138, Application US/09543681A
      ; Patent No. 6605709
      ; GENERAL INFORMATION:
      ; APPLICANT: GARY BRETON
      ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
      ; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
      ; FILE REFERENCE: 2709.1002-001
      ; CURRENT APPLICATION NUMBER: US/09/543, 681A
      ; PRIOR FILING DATE: 2000-04-05
      ; PRIOR FILING DATE: 1999-04-09
      ; NUMBER OF SEQ ID NOS: 8344
      ; SEQ ID NO 1138
      ; LENGTH: 1125
      ; TYPE: DNA
```

```
; ORGANISM: Proteus mirabilis
US-09-543-681A-1138
Alignment Scores:
Pred. No.: 5 71e-09
Score: 200.00
Percent Similarity: 46.18%
Best Local Similarity: 26.51%
Query Match: 12.23%
DB: 4 Gaps: 6
US-10-089-514-6 (1-322) x US-09-543-681A-1138 (1-1125)
QY      8 ValValAlaGlyGlySerGlyValAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
      304 GTCATTTGTCGGGGGGAACGGAAGAAATGCGGACACTTTTTCGCTCTTACACTTCT 363
      QY      28 GlySerArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
      364 GCCTATCAAGTC-----GAAAGCTTGA 387
      QY      48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
      388 GCGATGAGTGCAGAAAGTAAATCTCGCTATTTTGGC-----GATCCTGATG 438
      QY      68 ValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAlaProValThrArgLeu 87
      439 GTGATATACAGTCCCAATCCATTAAACGTTGATGTTATGACCAATTACGCCCTTG 498
      QY      88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly--MetAlaAla 106
      499 CCAGAA--AATTGTTTATGTTAGTATGACCTCAATCAACAGCCCTTTAGAGCG 555
      QY      107 GluLeuAlaAlaHisAlaProGlyValAlaGlnHisGlyLeuAsnProMetPheAlaPro 126
      556 AAGTTAAAGCAATATATGCCCCCTGTT-----TTAGGACTATCAATCAATGTTGATCA 609
      QY      127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThr---ArgAspGlyProGly 145
      610 GATGTGCCAAGTTTGAACAAACAGTTATCGCTTATTTGGAAGGCGAGATTTACACAT 669
      QY      146 ValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArgProValArgLeuThrAla 165
      670 TTTGAGTGTGCTTTCGACCACTTATGCTGTGGGAGCAAGAGTAGAGGCTATTAATGCA 729
      QY      166 GluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuSer 185
      730 CAAGACAGATAAATAATATGATTTTATCCAAAGCATTCGACATTTTACTTATTTGCA 789
      QY      186 PheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaAlaAlaThrAlaPro 205
      790 TATGGCAACATTTTGTGTAAGAAGATGCGATTTTATGACAGATTTTATGATTAATCTTCT 849
      QY      206 ProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlyGlySerProGluVal 225
      850 CCCATCTATGTTTATGAGCTTACGATGATGTTGATGCTCTTTCGTCAGATCCACATTA 909
      QY      226 TyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgArgAlaLeuAlaGlu 245
      910 TATGCGAATATTAATTTATCTTCACAGAAACATTAACCTTGAATCGTATATATATCAT 969
      QY      246 AlaLeuArgSerPheAlaAlaLeuVal 254
      970 TCATTATGGAAGCTATAGCTTATTA 996
      RESULT 11
      US-09-902-540-5637
      ; Sequence 5637, Application US/09902540
      ; Patent No. 6833447
      ; GENERAL INFORMATION:
      ; APPLICANT: Goldman, Barry S.
      ; APPLICANT: Hinkle, Gregory J.
      ; APPLICANT: Slater, Steven C.
```



```

; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO: 5637
; LENGTH: 1038
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-5637

Alignment Scores:
Pred. No.: 5,64e-06 Length: 1038
Score: 166.50 Matches: 86
Percent Similarity: 41.22% Conservative: 36
Best Local Similarity: 29.05% Mismatches: 144
Query Match: 10.18% Indels: 31
DB: 4 Gaps: 10

US-10-089-514-6 (1-322) x US-09-902-540-5637 (1-1038)
OY 7 ServaIvalValGlySerGlyValAlaValGlyMetPheAlaGlyLeuAlaArgGlu 26
DB 10 AGCATGCCGCTGCTGGGAGTACGCGCCCTTGGCCCTGCTGCTGCTGCTGCTGCTG 69
OY 27 AlAGlySerArgThrLeuValAlaAspLeuValProProGlyArg-----Pro 43
DB 70 GCGGCG-----ATCCCGCATCGGCTGTTGAGCGCGCGCATGACGATGCGCC 117
OY 44 AspAlaCysLeuValAlaAspValThrAlaProGlyProGluLeuAlaAlaLeuArg 63
DB 118 GACGCA-----CTGAGGCGCACCGCCTCGCGGAGAGTCTGAG 156
OY 64 AspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaAlaPro 83
DB 157 GCGCGCGGCTGTGTGTCTCTCATGCGCGCTCTGAAATGCGCTCGCTGAGAG 216
OY 84 ValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly 103
DB 217 CTTCGCGCGCGCTGTGCGCGCGCGAGAGCGTCTGACGCTGCGAAGCGTCAAGTGC 276
OY 104 MetAlaAlaGluLeuAlaAla--HisAlaProGlyValGlnHisValGlyLeuAlaPro 122
DB 277 CCGTGTGAGAGTGTGCGCTCTGCTGTGGGCGGAAATTCCTCGGTGGGACAGCATTC 336
OY 123 MetPheAlaProAlaAlaGlyMetThrArgProValAla-----AlaValAla 139
DB 337 CTGTTCGCGCGCGCGAGTCT--GGCGAGGCGCGACCTGCTGCGCGACGCTGCTGCC 395
OY 140 ThrArgAspGlyPro--Gly--ValThrAlaLeuLeuArgLeuValGluGlyValGly 158
DB 396 GAAAGAACTGCAACCGAGGCGCTGCGCAAGCGCGCGGCTCTTTCGAAGCATTCGCG 455
OY 158 ValArgProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAla 178
DB 456 TGAAGTGAAGAGAGTCTGCTGCGCGGATGCGACGCGCGTGAATGCTGCGACGATGCT 515
OY 178 ValThrAlaAlaValLeuLeuSerThrGlyLeuAlaLeuAlaArgLeuGlyValAlaAla 198
DB 516 CACC-----TTCTTCTGCGCGCGCGAGGTGCTCAAGCGGAGCGCGGAAAGACCTG-- 567
OY 198 GalAlaLeuAlaAlaThrAlaProProProHisGlnValLeuLeuAlaLeuValAlaArg 218
DB 568 -----CCTTTCGCGCGCGCGCGAGTCTCAAGCGCGCTGCGCTGAGAGAGTACGC 617
OY 218 IleuGlyGlySerProGluValArgGlyAspIleGlnArgSerAsnProArgAlaAla 238
DB 618 ACCGCTGAGAGTCCCATCTTCTTCGCGCGTGTGCGAGTCCGAGAAATCCCTACCGCGGA 677
OY 238 ValArgAspGlyPro--Gly--ValThrAlaLeuLeuArgLeuValGluGlyValAsp 258

```

```

DB 678 CGAGCGCTGCGCGCTGCTGAGAGCGCTTCAAGCATTCAC-----CTGGCGTGTATTC 721
OY 258 AspArgAlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyValCysAsp 278
DB 732 GCGCGCGCTGCGCGCGCGCGAGCTGTGCGCGCGCG-----CCGCGCTGACGGAAGT 785
OY 278 ValAGlyAsnLeuAspGlyValPheGlyLeuLeuArgLeuMet 293
DB 786 TCGGAGTGTGTGATGCTGTGACCGGTGACCTGTTCAACTCTCTG 831

RESULT 12
US-09-902-540-1269/c
; Sequence 1269, Application US/09902540
; Patent No. 683347
; GENERAL INFORMATION:
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Wiegand, Roger C.
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof
; FILE REFERENCE: 38-10(15849)B
; CURRENT APPLICATION NUMBER: US/09/902,540
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: 60/217,883
; PRIOR FILING DATE: 2000-07-10
; NUMBER OF SEQ ID NOS: 16825
; SEQ ID NO: 1269
; LENGTH: 49225
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-09-902-540-1269

Alignment Scores:
Pred. No.: 0.00121 Length: 49225
Score: 166.50 Matches: 86
Percent Similarity: 41.22% Conservative: 36
Best Local Similarity: 29.05% Mismatches: 144
Query Match: 10.18% Indels: 31
DB: 4 Gaps: 10

US-10-089-514-6 (1-322) x US-09-902-540-1269 (1-49225)
OY 7 ServaIvalValGlySerGlyValAlaValGlyMetPheAlaGlyLeuAlaArgGlu 26
DB 41669 AGCATGCCGCTGCTGGGAGTACGCGCCCTTGGCCCTGCTGCTGCTGCTGCTGCTG 41610
OY 27 AlAGlySerArgThrLeuValAlaAspLeuValProProGlyArg-----Pro 43
DB 41609 GCGGCG-----ATCCCGCATCGGCTGTTGAGCGCGCGCATGACGATGACGCTGCC 41562
OY 44 AspAlaCysLeuValAlaAspValThrAlaProGlyProGluLeuAlaAlaLeuArg 63
DB 41561 GACGCA-----CTGAGGCGCACCGCCTCGCGGAGAGTCTGAG 41523
OY 64 AspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaAlaPro 83
DB 41522 GCGCGCGGCTGTGTGTCTCTCATGCGCGCTGTGAAATGCGCTGCTGCGAGAG 41463
OY 84 ValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly 103
DB 41462 CTTCGCGCGCGCTGTGCGCGCGCGAGAGTCTGACGCTGCGAAGCGCTCAAGTGC 41403
OY 104 MetAlaAlaGluLeuAlaAla--HisAlaProGlyValGlnHisValGlyLeuAlaPro 122
DB 41402 CCGTGTGAGAGTGTGCGCTCCGTTGGCGGAGCATTCCTCGGTGGGACAGCATTC 41343
OY 123 MetPheAlaProAlaAlaGlyMetThrArgProValAla-----AlaValAla 139
DB 41342 CTGTTCGCGCGCGCGAGTCT--GGCGAGGCGCACCTGCTGCGCGAGTGTGTGCC 41284
OY 140 ThrArgAspGlyPro--Gly--ValThrAlaLeuLeuArgLeuValGluGlyValGly 158

```



```

: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 31
: LENGTH: 1362
: TYPE: DNA
: ORGANISM: Bacteria
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(1362)
US-09-724-797-31

Alignment Scores:
Pred. No.: 0.000234 Length: 1362
Score: 150.50 Matches: 101
Percent Similarity: 35.03% Conservatvie: 30
Best Local Similarity: 27.01% Mismatches: 136
Query Match: 9.20% Indels: 108
DB: 4 Gaps: 18

US-10-089-514-6 (1-322) x US-09-724-797-31 (1-1362)

Oy 5 ProaGSeValaIyValaIyGlySerGlyAlaValaIyGlyMetPheAlaGlyLeuLeu 24
Db 34 CCGGTGAAGGTGTCATCGCCGGCCGGCGGCTACGTGGCACTGTCTCGCGTCACCCCTC 93
Oy 25 ArgGluAlaGlySerArgThrLeuValAlaAspLeuValProProProGlyArgProAsp 44
Db 94 GCCGGCCGGCGGGCCGAGGTGTGTCCGGTCGAC-----AGCAGCCCGGGCACCGTCGC 147
Oy 45 AlAcYbLeuValGlyAspValThrAlaProGlyProGlyLeuValAlaAlaLeuArgAsp 64
Db 148 GACCTGCGGGCCGGCCGGTCGGCTGGCTCCGAGCCCGGCTGGCCGGCCGCTCCGGGAC 207
Oy 65 -----AlaAsp 66
Db 208 CTGCGCGGACCGGACGCTGACGCGGAGACGTCGACGCCGGTCGGCGGGCGGAC 267
Oy 67 LeuValLeuLeuAlaVal-----HisGluAspValAlaLeuLys 79
Db 268 GTGGTGATGTGTCACGGTCCGACACCCGACCGACCCGGCCACGATGTGTCACCGACAG 327
Oy 80 AlaValAlaProValThrArgLeu-----MetArgProGlyAlaLeuLeuAlaAsp 96
Db 328 CTCGTGCGGGCGGTGCGACGAGATCGCCCGCGGCTGGCGCCGGGCACTGTGTATCTC 387
Oy 97 ThrLeuSerValArgThrGlyMetAlaAlaGlyLeuAlaAla-----HisAlaPro 113
Db 388 AAGTGAAGCGTCTCCCGGGGACACACCCGAGACCTGTGTGGCGCCCTGCTGAGAGAGCGG 447
Oy 114 GlyAlaGlnHis-----ValGlyLeuAsnProMetPheAlaProAlaAlaGlyMet 130
Db 448 GGGCTGTGTGTCACGAGCGGACATCTGGGGCTG-----GCCTTGTGCGGAGCGGCTGCC 501
Oy 131 ThrGlyArgProValAlaAlaValAlaValThrArgAspGlyProGlyValThrAlaLeuLeu 150
Db 502 GAGGGAGTGGCCCTGTCGGAGGTGGCGAGCG-CTG 534
Oy 151 ArgLeuValaIyGlyValaIyGlyArgProValArgLeuThrAlaGluGlnHisAspArg 170
Db 535 CCGGTGTGTGTGGGTGGGTGGCGC-----CCGCGC 564
Oy 171 ThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuSerPheGlyLeuAlaLeu 190
Db 565 AGCGCGCGCGGCGGACGATCTCGGCGGTCCGC----- 600
Oy 191 AlaArgLeuGlyValaAspValArgAlaLeuAlaAlaThr-AlaProProProHisGlnVa 210
Db 601 -----CTCGGGGTGTCAGCTCCGGCAGGTGCCCTCGCCGAGTGTCCGCCGAGGTGTCA-- 652
Oy 210 IleuLeuAlaLeuLeuAlaArgValaLeuGlyGlySerProGlyValaIyThrGlyAspIleG 230
Db 653 -----AGCTCCGACCAACTGTGTGATGACCGCAACGT--GGCATGCGCA 697
Oy 230 narGerAenPro-----ArgAlaAlaSerAlaArgArgAlaLeuAlaGluAlaLeuAr 248

```

```

Db      698  ACDAACTCGCCCGGTACTCGCGGGGTCTGGGGGTGAACGTCTCTGCAGCTGATCGGCC-- 755
Qy      248  gSerPheAlaLaLeuValGlyAspAspProAspArgAlaAspAlaProGlyArgAlaAs 266
Db      756  -----GGCGAACACCTTGCCCAAGGACAGACGATGGTGAACCTGC 796
Qy      268  pAlaProGlyHisProGlyGlyCys-----AspGlyAlaGlyHisLeuAspG 284
Db      797  TGTGTGCGCGG-----GGTGGGTGTGGCGCGCTCTCTGTCGACGAAGAACCCGTGATGG 855
Qy      284  yAlaPheGlyGlyLeuLeuArgArgLeuMetGlyProGlyLeuAla----- 298
Db      851  CGTGGCGGGAACG-----CGGGACCGGGGGGTCTCCCTCGCGACGCTCGAGA 899
Qy      299  -----AlaGlyGlnAspHisCysGlnGlyLeuPheArgThrLeuHisArg----- 313
Db      899  CGGCGCGCGGGGTCAACGACGACATGCCCGCGACACCGCGCGGTATCGCCGACGAGC 955
Qy      314  -----ThrAspAspGlyGlyGlyLeuAspArg 322
Db      959  TGTCAAGCTGGGACGGGATCGGAACGACAGACGATGCG 998

RESULT 15
US-09-589-892B-3/c
; Sequence 3, Application US/09589892B
; Patent No. 6689583
; GENERAL INFORMATION:
; APPLICANT: Jennewein, Thomas
; APPLICANT: Laible, Gortz
; APPLICANT: O'Carroll, Donal
; APPLICANT: Eisenhaber, Frank
; TITLE OF INVENTION: Chromatin-Regulator Genes
; FILE REFERENCE: 0652.1670001
; CURRENT APPLICATION NUMBER: US/09/589,892B
; CURRENT FILING DATE: 2000-06-09
; PRIOR APPLICATION NUMBER: US 08/945,988
; PRIOR FILING DATE: 1997-11-10
; PRIOR APPLICATION NUMBER: PCT/EP96/01818
; PRIOR FILING DATE: 1996-05-02
; PRIOR APPLICATION NUMBER: DE 195 16 776.7
; PRIOR FILING DATE: 1995-05-10
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 2732
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(44)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (45)..(1283)
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: (1284)..(2732)
US-09-589-892B-3

Alignment Scores:
Pred. No.: 0.00844 Length: 2732
Score: 149.00 Matches: 82
Percent Similarity: 44.96% Conservative: 34
Best Local Similarity: 31.78% Mismatches: 97
Query Match: 9.11% Indels: 45
DB: Gaps: 14

US-10-089-514-6 (1-322) x US-09-589-892B-3 (1-2732)
Qy      9  ValValGlyGlySerGlyValAlaGlyGlyMetPheAlaGlyLeuLeuArgGlyAlaGly 28
Db      989  ATGTGTGGCGGCGATCCACGGTGTAAACGCTCTCCACGTAGTCCAGGTCAAGAGAGTAGGT 930

```


GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:57:48 ; Search time 708.632 Seconds
(without alignments)
3137.453 Million cell updates/sec

Title: US-10-089-514-6

Perfect score: 1635
Sequence: 1 MSGPFRSVVVGSGAVGMF.....HCELPFTLRLTDEGDKR 322

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 7442561 seqs, 3452328358 residues

Total number of hits satisfying chosen parameters: 14885122

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODE=frame+ p2n.model -DEV=X1h
-Q=/cgn2_1/USPTO.spool.h/US10089514/runat_04102005.105745.8055/app_query.fasta_1.1621
-DB=Published Applications NA -QFMT=fastap -SUFFIX=emb -MINMATCH=0.1
-LOOPC=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blonsum62
-TRANS=human40.ccl -LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THE MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MATEL=200000000 -USER=US10089514@cgn_1.1.920@runat_04102005.105745.8055
-NCPU=6 -ICPU=3 -NO MAP -LARGQUERY -NEG SCORES=0 -WAIT -DSBLOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Published Applications_NA:*
1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:*
4: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:*
5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/prodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
17: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
18: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
19: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
20: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
21: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:*
22: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:*
23: /cgn2_6/prodata/1/pubpna/US11_PUBCOMB.seq:*
24: /cgn2_6/prodata/1/pubpna/US11_NEW_PUB.seq:*
25: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:*
26: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1635	100.0	969	20	US-10-472-587-5
2	1588	97.1	3305	14	US-10-126-927-68
3	1588	97.1	12391	14	US-10-126-927-67
4	237	14.5	1122	16	US-10-137-310-3
5	233	14.3	1122	19	US-10-433-556-18
6	233	14.3	1238	16	US-10-380-133-7
7	225.5	13.8	1122	16	US-10-137-310-1
8	225.5	13.8	1122	18	US-10-634-548-16
9	225.5	13.8	1152	17	US-10-433-556-20
10	216	13.2	1830121	19	US-10-329-670-1
11	216	13.2	1830121	20	US-10-158-865-1
12	216	13.2	1830121	22	US-10-981-687-1
13	213	13.0	1792	22	US-10-450-763-1605
14	174.5	10.7	2731748	19	US-10-297-465A-1
15	165	10.1	2025	17	US-10-369-493-37621
16	159	9.7	88421	9	US-09-976-059-1
17	153.5	9.4	954	19	US-10-437-963-101181
18	152	9.3	891	19	US-10-375-039-13
19	151	9.2	9025608	15	US-10-156-761-1
20	150.5	9.2	1491	15	US-10-156-761-1868
21	150.5	9.2	36602	21	US-10-762-107-1
22	150.5	9.2	9025608	15	US-10-156-761-1
23	150	9.2	42291	21	US-10-485-710-4
24	149	9.1	1236	21	US-10-500-416-2
25	149	9.1	1023	21	US-10-773-302-3
26	148	9.1	1023	22	US-10-898-775-27
27	147.5	9.1	1023	22	US-10-911-160-21
28	147.5	9.0	1220	19	US-10-437-963-86432
29	147.5	9.0	1212	11	US-09-758-759-182
30	147.5	9.0	37116	17	US-10-107-431-279
31	146	8.9	2120	22	US-10-499-065A-239
32	146	8.9	2280	22	US-10-499-065A-242
33	146	8.9	2431	22	US-10-499-065A-244
34	146	8.9	2457	22	US-10-499-065A-238
35	146	8.9	2489	22	US-10-499-065A-237
36	146	8.9	2520	22	US-10-499-065A-236
37	146	8.9	2641	22	US-10-499-065A-243
38	146	8.9	18435	15	US-10-156-761-412
39	146	8.9	100000	15	US-10-156-761-15103
40	145	8.9	3957	16	US-10-200-562-193
41	145	8.9	3957	16	US-10-237-451-193
42	145	8.9	3957	22	US-10-945-050-193
43	145	8.9	154746	10	US-09-827-688-8
44	145	8.9	154746	10	US-09-827-688-8
45	143.5	8.8	25617	21	US-10-485-710-7

ALIGNMENTS

RESULT 1
US-10-472-587-5
; Sequence 5, Application US/10472587
; Publication No. US20040214274A1
; GENERAL INFORMATION:
; APPLICANT: YANAI, Koji
; APPLICANT: SUMIDA, Naomi
; APPLICANT: MATANABE, Manabu
; APPLICANT: MORIYA, Tatsuki
; APPLICANT: MURAKAMI, Takeshi
; TITLE OF INVENTION: Transformants Producing Substance P1022 Derivatives, Methods for
; FILE REFERENCE: 2003-1302A/MMC/00144
; CURRENT APPLICATION NUMBER: US/10/472,587
; CURRENT FILING DATE: 2003-09-22
; PRIOR APPLICATION NUMBER: 82227/2001

[illegible]

QY	261	AlaAspAlaProGlyValArgAlaAspAlaProGlyHisProGlyValGlyAspGlyValGly	280
Db	781	GCCAGCGCCCCGGGACGGCCGACGCCCCCGCCATCCCGGGAGTGCAGCGGCGCGG	840
QY	281	AsnLeuAspGlyValIlePheGlyGlyLeuAspArgLeuMetGlyProGlyLeuValAlaGly	300
Db	841	AACTTCAGCGCGCTTCGCGGAACTCGCGCGGCTCATGGAGCCGGAGCTCGCGCGGCG	900
QY	301	GlnAspHisCysGlnGlyLeuPheArgThreonHisArgThrAspAspGlyGlyValys	320
Db	901	CAGAGCACTGCACGAGAGCTGTTCGACCTTCACGACCAACGACGAGAGGAGAG	960
QY	321	AspArg 322	
Db	961	GACCGA 966	
RESULT 2			
US-10-126-927-68			
; Sequence 68, Application US/10126927			
; Publication No. US20030082575A1			
GENERAL INFORMATION:			
; APPLICANT: The Scripps Research Institute			
; APPLICANT: Schultz, Peter G			
; APPLICANT: Wang, Lei			
; APPLICANT: Anderson, John C			
; APPLICANT: Chin, Jason			
; APPLICANT: Liu, David R			
; APPLICANT: Magliery, Thomas			
; APPLICANT: Meggers, Eric L			
; APPLICANT: Mehl, Ryan A			
; APPLICANT: Pastinak, Miro			
; APPLICANT: Santoro, Stephen W			
; APPLICANT: Zhang, Zhiwen			
; TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids			
; FILE REFERENCE: 54-000120US			
; CURRENT APPLICATION NUMBER: US/10/126, 927			
; CURRENT FILING DATE: 2002-04-19			
; PRIOR APPLICATION NUMBER: US 60/285, 020			
; PRIOR FILING DATE: 2001-04-19			
; PRIOR APPLICATION NUMBER: US 60/355, 514			
; PRIOR FILING DATE: 2002-02-06			
; NUMBER OF SEQ ID NOS: 79			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 68			
; LENGTH: 3305			
; TYPE: DNA			
; ORGANISM: Streptomyces venezuelae			
US-10-126-927-68			
Alignment Scores:			
Pred. No.: 4.64e-153 Length: 3305			
Score: 1588.00 Matches: 315			
Percent Similarity: 96.65% Conservative: 2			
Best Local Similarity: 96.04% Mismatches: 5			
Query Match: 97.13% Indels: 6			
DB: 14 Gaps: 1			
US-10-089-514-6 (1-322) x US-10-126-927-68 (1-3305)			
QY	1	MetSerGlyPheProArgSerValValAlaGlyGlySerGlyAlaValGlyGlyMetPhe	20
Db	2007	ATGAGCGGCTTCCCCCGGAGGCTGCTGCTGCGGCGGAGCGGAGGCGGATGTC	2066
QY	21	AlaGlyLeuLeuAspGlyAlaGlySerArgThrLeuValValAspLeuValProProPro	40
Db	2067	GCCGGGCTGCTCGCGGAGGCGGCGGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTG	2126
QY	41	GlyArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGlyLeuValAla	60
Db	2127	GGAAGGCGCGGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2186
QY	61	AlaLeuArgAspAlaAspLeuValLeuLeuValAlaHisGlyAspValAlaLeuValAla	80

Db	2187	GCCTCCGGGACGGGACCTGCTGCTGCTGCCCGTACACGAGACGTGGCCCTCAAGGCC	2246
Qy	81	ValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal	100
Db	2247	GTGGCGCCCGGTGACCCCGGCTCATGGACCGGGCGGGCTGCTCCGCGACACCTGTGCGCTG	2306
Qy	101	ArgThrGlyMetLeuAlaGlyLeuLeuAlaAlaIleAlaProGlyValAlaGlnIleValGlyLeu	120
Db	2307	CGAGCGGGCAGATGGCCGGGAGCTCGCGGCCACAGCCCGCCGGCGTCCAGACAGTGGGCTTC	2366
Qy	121	AsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaValValThr	140
Db	2367	AACCGAGTTCGCCCCCGCGCGCGGATGACCGGCGGCCCCGTGGCCGCGCTGTACAC	2426
Qy	141	ArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyValArgPro	160
Db	2427	AGGACCGGGCGGGCGTCACGGCGCTGCTGCGGCTGCTCGAGGGGGGGCGGACGAGGCC	2486
Qy	161	ValArgLeuThrAlaGlyGlnIleAspArgThrThrAlaAlaThrGlnAlaLeuThrHis	180
Db	2487	GTACGGCTCAGCGGGGAGAGACAGACCGGACGACGGGGGAGACCAAGCCCTTACGCA	2546
Qy	181	AlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeu	200
Db	2547	GCCTGTACCTCTCTCTCGGGCTGCGCTCGCCCGCTCGGCGTTCAGAGTCCGGGCTTG	2606
Qy	201	AlaAlaThrAlaProProThrIleGlnValLeuLeuAlaLeuLeuAlaArgValLeuGly	220
Db	2607	GGCGGACGCGACCGCGCGCCCAACAGGTGCTGCTCGGCTCTGAGCCCGGTGCTGGC	2666
Qy	221	GlySerProGlyValTyrGlyAspTyrGlnArgSerAsnProArgAlaIleSerAlaArg	240
Db	2667	GGCAGCCCGGAGTGTACGGGGAGATCCAGCGGTCCAAACCCCGGGCGGCGTCCGCGCGC	2726
Qy	241	ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAsp-----	256
Db	2727	CGGGGCTCGCGGAGGCGCTGCGCTCTGCGCGGCTGATCGGCGAGACCCGGACCGC	2786
Qy	257	-----AspProAspArgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly	274
Db	2787	GCGGAGGACCGCGGACCGCGCGGACGACCCGACCGGACCGGACCGGCGGCGG	2846
Qy	275	GlyCyAspArgGlyAlaGlyAsnLeuAspGlyValPheGlyGlyLeuArgArgLeuMetGly	294
Db	2847	GGAATGCCACGGCGCGGGGAACTTGACGCGCTTTCGAGAACTCCGCGGCTCATGGGA	2906
Qy	295	ProGlyLeuAlaAlaGlyGlnAspHisSerGlnGlyLeuPheArgThrLeuHisArgThr	314
Db	2907	CCGGAAGCTCGGGGGGCGGACGACCACTGCGCAGAGAGCTGTTCCGACCCCTTCACCGCAC	2966
Qy	315	AspAspGlyGlyGlyValAspArg 322	
Db	2967	GACGACGAGCGGACGAGAGACCGA 2990	
RESULT 3			
US-10-126-927-67			
: Sequence 67, Application US/10126927			
: Publication No. US20030082575A1			
GENERAL INFORMATION:			
APPLICANT: The Scripps Research Institute			
APPLICANT: Schultz, Peter G			
APPLICANT: Wang, Lei			
APPLICANT: Anderson, John C			
APPLICANT: Chin, Jason			
APPLICANT: Liu, David R			
APPLICANT: Magliery, Thomas			
APPLICANT: Meggers, Eric L			
APPLICANT: Mehl, Ryan A			
APPLICANT: Pastinak, Miro			
APPLICANT: Santoro, Stephen W			
APPLICANT: Zhang, Zhiwen			
: TITLE OF INVENTION: In Vivo Incorporation of Unnatural Amino Acids			

```

FILE REFERENCE: 54-000120US
CURRENT APPLICATION NUMBER: US/10/126,927;
CURRENT FILING DATE: 2002-04-19
PRIOR APPLICATION NUMBER: US 60/285,030
PRIOR FILING DATE: 2001-04-19
PRIOR APPLICATION NUMBER: US 60/355,514
PRIOR FILING DATE: 2002-02-06
NUMBER OF SEQ ID NOS: 79
SOFTWARE: PatentIn version 3.1
SEQ ID NO 67
LENGTH: 12391
TYPE: DNA
ORGANISM: Plasmiid pSC101, Streptomyces venezuelae papABC
US-10-126-927-67

Alignment Scores:
Pred. No.:          2,12e-152          Length:          12391
Score:              1588.00             Matches:         315
Percent Similarity: 96.65%             Conservative:    2
Best Local Similarity: 96.04%           Mismatches:     5
Query Match:        97.13%              Indels:         6
DB:                 14                   Gaps:           1

```

US-10-089-514-6-(1-322)	x	US-10-126-927-67	(1-12391)
QY	1	MecSerGclYheProAagSerValValValGlyGlySerGlyAlaValGlyGlyMecPhe	20
Db	2361	ATGATGTGGCTTTCCTCCCGAGACGTCGTGTGGCGGCGAGCGAGCGATGGCTTC	2420
QY	21	AlGglYleuMuyrGglValIagIySerYarGThrIeuValValAapIleuValProProPro	40
Db	2421	GCCGGGCTGCTGGCGGAGGGGGGAGCGGACGCTGTCGTGCACCTGTACCCCGCCG	2480
QY	41	GlyYarGProAaPaIaCyLeuValGlyAaPValThraIaProGlyProGluIleuAlaIa	60
Db	2481	GAGACGGCGGACCGCTGCTGTGTGGGAGATCACCGCGCGGGCGCCGAGCTCGCCG	2540
QY	61	AlaIeuYarGaaPaIaAaPleuValIleuIleuAlaIhIsgIuAaPValAlaIeuIySaIa	80
Db	2541	GCCCTCCGGGAGCGGAGACTGTCGTCTGCTCGCTACAGAGAGCGTGTGCTCAAGGCG	2600
QY	81	ValAlaProValThraGleuMecYarGProGlyAlaIeuIleuAlaAapThrIleuSerVal	100
Db	2601	GTCGCGCCGCTGACCCGCTCATATGCACCGAGCGCGCTGCTGCGGACACCTTTCGCTC	2660
QY	101	ArgThrGlyMecAlaIagIleuIleuAlaIaIhIsgIaProGlyValGlnHIsValGlyIleu	120
Db	2661	CGAGCGGGCATGGCGCGGAGACTGCGGGCCACGCGCCCGCGCTCCAGACATGGGCTTC	2720
QY	121	AsnProMecPheAlaProAlaIagIyMecThrGlyYarGProValAlaIaValValThr	140
Db	2721	AACCGATGTTCCGCGCGCGCGCGGCGGATACCGGCGCGCGCGCGCTGTGTACCC	2780
QY	141	ArgAaPglyProGlyValThraIaIeuIleuYarGleuValGlyGlyGlyYarGPro	160
Db	2781	AGGACCGGGCGCGGCGCTCACAGCGCTGCTCGCTCCAGAGGCGCGCGGCGAGCGCC	2840
QY	161	ValArgIeuThraIagIugIuHIsAaPaYrGThrThraIaIaThrGlnAlaIeuThrHis	180
Db	2841	GTCAGGCTTCAGCGCGGAGGAGACAGACCGGACGAGCGGGCGAGCCACAGGCTTACGGAC	2900
QY	181	AlaValIleuIeuSerPheGlyIleuAlaIeuAlaArgIeuGlyValAaPValArgAlaIeu	200
Db	2901	GCGGTATCTCTTCCTTCGGGAGTCGCGCTCGCGCGCTCGGGCGTCAAGTCGGGCGCTG	2960
QY	201	AlaAlaThraIaProProProHIsGlnValIleuIleuAlaIeuIleuAlaArgValIeuGly	220
Db	2961	GCGCGGACGCGACCGCGCGCCACAGAGTCTGCTCCCTCGCGCGCGGTGTCTGGCG	3020
QY	221	GlySerProGluValYrGlyAaPglyIeGlnArgSerAsnProArgAlaIaSerAlaArg	240
Db	3021	GCGAGCGCCGAGGTGATCGGGGACATCCACAGGCTCAACCCCGGGGGGGGTTCGCGCGC	3080


```
QY 241 ATGAlaleuAlaGluAlaleuArgSerPheAlaAlaleuValGlyAsp----- 256
Db 3081 CGGGCCCTCGCCAGAGCCCTCGCTCTTCCGCCGCTGATCGGCACGACCCGACCCG 3140
QY 257 -----AspProspaRgAlaAspAlaProGlyArgAlaAspAlaProGlyHisProGly 274
Db 3141 GCCAGAGACCCGGAGCCGCGCCGAGACCCGACCCGACCCGACCAACCCGCGCATCCCGGG 3200
QY 275 GlyCyAspGlyValaGlyAsnLeuAspGlyValaPheGlyGluLeuArgArgLeuMetGly 294
Db 3201 GATGAGGAGCGGCGCCGGAAACCTCGACGGGCTTTGAGAGAACTCCGCGGCTCATGCGA 3260
QY 295 ProGluLeuAlaAlaGlyGlnAspHisGlyGlnGluLeuPheArgThrLeuHisArgThr 314
Db 3261 CCGAGACTCGCGCGCGCCGACGACCACTGCGCAGAGAGCTGTTCCGCACTCCACCCGACCC 3320
QY 315 AspaRgGluGlyGlyLysAspArg 322
Db 3321 GACGACGAAAGCGAGAGACCGA 3344

RESULT 4
US-10-137-310-3
; Sequence 3. Application US/10137310
; Publication No. US20030176675A1
; GENERAL INFORMATION:
; APPLICANT: Valentim, Henry E.
; APPLICANT: Mitsky, Timothy A.
; TITLE OF INVENTION: Tyra Genes and Uses Thereof
; FILE REFERENCE: 16515.147
; CURRENT APPLICATION NUMBER: US/10/137, 310
; CURRENT FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 60/289,527
; PRIOR FILING DATE: 2001-05-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
US-10-137-310-3

Alignment Scores:
Pred. No.: 1,32e-14 Length: 1122
Score: 237.00 Matches: 78
Percent Similarity: 45.39% Conservative: 50
Best Local Similarity: 27.66% Mismatches: 110
Query Match: 14,50% Indels: 44
DB: 16 Gaps: 8

US-10-089-514-6 (1-322) x US-10-137-310-3 (1-1122)
QY 8 ValValValGlyGlySerGlyAlaValaGlyGlyMetPheAlaGlyLeuArgGluAla 27
Db 304 GTTATGCTTGGCGCGCGCGCTCAGATGGAGCGTCTGTGTCAGAGAAAGCTGACACTCTCG 363
QY 28 GlySerArgThrLeuValaAspLeuValaProProGlyArgProAspAlaCysLeu 47
Db 364 GGTATCAGGTGGCGGATTCGAG-----CAACATGACTGGATTCGA 405
QY 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
Db 406 GCGGCTGATGTTGTTGCC-----GATGCCGAATG 435
QY 68 ValLeuLeuAlaValHisGluAspValaAlaLeuGlyAlaValaProValThrArgLeu 87
Db 436 GTGATGTAGTGTGCCAATCCAGTTACTGACCAAGTTATGGCAAAATTACCGCCTTTA 495
QY 88 MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly--MetAlaAla 106
Db 496 CGAAG--GATGTATCTCTGTTGATCGGCACTCAGTGAATAATGACCATTCACGGCC 552
QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
```

```
Db 553 ATGTCGCGCGCGCAGCAGTGGCCCGTA-----CTGGGGTTTACACCAATGTTGGTCCG 606
QY 127 AlaAlaGlyMetThrGlyArgProValaAlaAlaValaThrArgAspGlyProGlyVal 146
Db 607 GACAGCGGTAGCTCGCAAGCAAGT-----GTGGTGTGGTGTATGAGACTTAACCG 660
QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValaArgLeu 163
Db 661 GAAGCATCAACCAATGTTCTTGAGCAAAATTTCAGTCTGGGGCGCTTGCAATCGTATT 720
QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
Db 721 AGGCGCGTGCACACATCAATATGCGTTTATTCAGGCACCTGGCCACTTGGTACT 780
QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValaAspValArgAlaLeuAlaThr 203
Db 781 TTTCCTTAGCGGCTGCACCTGGCAGAAAGAAATTTCACTTGAAGCACTTCTGGGCTTC 840
QY 204 AlaProProHisGlnValLeuLeuAlaLeuAlaArgValLeuGlySerPro 223
Db 841 TCTTCGCGATTTTACCGCCTTGAGCTGGCGATGTCGGGCGACTGTTCCGTCAGATCCG 900
QY 224 GluValTyrGlyAspGlnArgSerAsnProArgAlaAlaSerAla----- 239
Db 901 CAGCTTATGCCGACATTTATATGTCGACAGCGTTAATCTGCAATCAACGTTAC 960
QY 240 -----ArgArgAlaLeuAla 244
Db 961 TATAAGCTTTGGCGCAGCGCATTAAGTTGCTGAGCAGGCGGATTAAGCGCTTATT 1020
QY 245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
Db 1021 GACAGTTCCGCAAGGTGAGACACTGTCGCGGATTAAGCAACAGCTTTTCAAGTGA 1080
QY 265 GlyArg 266
Db 1081 AGCCGC 1086

RESULT 5
US-10-433-556-18
; Sequence 18. Application US/10433556
; Publication No. US20040117872A1
; GENERAL INFORMATION:
; APPLICANT: Matrinege, Michel
; APPLICANT: Ripper, Pascal
; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
; FILE REFERENCE: 5500*120
; CURRENT APPLICATION NUMBER: US/10/433,556
; CURRENT FILING DATE: 2003-06-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 18
; LENGTH: 1122
; TYPE: DNA
; ORGANISM: Escherichia coli
; NAME/KEY: CDS
; LOCATION: (1)..(1122)
US-10-433-556-18

Alignment Scores:
Pred. No.: 3,4e-14 Length: 1122
Score: 233.00 Matches: 77
Percent Similarity: 45.04% Conservative: 50
Best Local Similarity: 27.30% Mismatches: 111
Query Match: 14,25% Indels: 44
DB: 19 Gaps: 8

US-10-089-514-6 (1-322) x US-10-433-556-18 (1-1122)
QY 8 ValValValGlyGlySerGlyAlaValaGlyGlyMetPheAlaGlyLeuArgGluAla 27
```

```

Db      304 GTTATCGTCGCGGTCGAGTCAGATGAGCAGCTGTTCCAGAAATGCTGACCCCTTCG 363
Qy      28  GlysArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
Db      364 GGTATTCAGGTGGGATTCGAG-----CAACATGACCTGGATCGA 405
Qy      48  ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
Db      406 GCGGCTGATTTGTGTC-----GATGCCGGAATG 435
Qy      68  ValLeuLeuAlaValHisGluAspValAlaLeuValAlaAlaProValThrArgLeu 87
Db      436 GTGATTGTTAGTGGCAATCCACGTACGTAGCAATGGAATAATGGCAATTAACGCCCTTTA 495
Qy      88  MetArgProGlyAlaLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
Db      496 CCGAAA---GATTGTATTCGTGTCGATCTGGCATCAGTGAATAATGGCCATTACAGGCC 552
Qy      107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
Db      553 ATGCTGTCGCGCATGATGTCGCGTG-----CTGGGCTTACACCGATCTTCGTCGC 606
Qy      127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Db      607 GACAGCGGTAGCTGCGCAAGCAAGTT-----GTGCTCGTGTGATGACGTAACCG 660
Qy      147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Db      661 GAAGCATACCAATGTTCTGAGAGCAATTCAGGTCTGGGCGCTCGGTCGATCGTATT 720
Qy      164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
Db      721 AGCGCCGTGAGCAGCATCGAATATGCGCTTATTACGGCATCTGCCCATCTTGCTACT 780
Qy      184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
Db      781 TTTCCTTACGGGCTGACCTGGCAGAAAGAAATGTCAGCTGACCACTTTCGGGCTC 840
Qy      204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerPro 223
Db      841 TCTTCGCGATTACCCCTTGGAGCTGCGATGTCGCGGCACTGTTGCTCAGATCCG 900
Qy      224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaLeuAla----- 239
Db      901 CACCTTATCCGACATCATTAATGTCGTACAGCGTAATCTGGGTTAATCAACGTTAC 960
Qy      240 -----ArgArgAlaLeuAla 244
Db      961 TATAAGGCTTTCGCGAGGCGATTGAGTTCGAGCAGGCGCATAGCAGGCGTTTATT 1020
Qy      245 GluAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaPro 264
Db      1021 GACAGTTTCCGCAAGGTGAGACACTGGTTCGCGATTACGACAGCGTTTTCAGAGTAA 1080
Qy      265 GlyArg 266
Db      1081 AGCCGC 1086

```

```

; ORGANISM: Escherichia coli
;
; FEATURE:
; NAME/KEY: gene
; LOCATION: (7)..(1232)
; OTHER INFORMATION: LysA gene coding for bifunctional chorismate
; OTHER INFORMATION: mutase / prephenate dehydrogenase
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (25)..(1143)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(6)
; OTHER INFORMATION: restriction site linker
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1233)..(1238)
; OTHER INFORMATION: restriction site linker
; US-10-380-132-7

Alignment Scores:
Pred. No.:      3,81e-14      Length:      1238
Score:          233.00      Matches:      77
Percent Similarity: 45.04%      Conservative: 50
Best Local Similarity: 27.30%      Mismatches: 111
Query Match:      14.25%      Indels:      44
DB:              16      Gaps:      8

US-10-089-514-6 (1-322) x US-10-380-132-7 (1-1238)
Qy      8  ValValValGlyGlySerGlyAlaValAlaGlyMetPheAlaGlyLeuLeuArgGluAla 27
Db      328 GTTATCGTCGCGGTCGCGTCAGATGAGACCGCTTTCGAGAAATGCTGACCCCTTCG 387
Qy      28  GlysArgThrLeuValValAspLeuValProProProGlyArgProAspAlaCysLeu 47
Db      388 GGTATTCAGGTGGGATTCGAG-----CAACATGACCTGGATCGA 429
Qy      48  ValGlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeu 67
Db      430 GCGGCTGATTAATTTGTC-----GATGCCGGAATG 459
Qy      68  ValLeuLeuAlaAlaHisGluAspValAlaLeuValAlaAlaProValThrArgLeu 87
Db      460 GTGATTGTTAGTGGCAATCCACGTACTGAGCAAGTTATTGGCAATTAACGCCCTTTA 519
Qy      88  MetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGly---MetAlaAla 106
Db      520 CCGAAA---GATTGTATTCGTGTCGATCTGGCATCAGTGAATAATGGCCATTACAGGCC 576
Qy      107 GluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuAsnProMetPheAlaPro 126
Db      577 ATGCTGTCGCGCATGATGTCGCGTG-----CTGGGCTTACACCGATGTCGTCGC 630
Qy      127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAspGlyProGlyVal 146
Db      631 GACAGCGGTAGCTGCGCAAGCAAGTT-----GTGCTCGTGTATGACGTAACCG 664
Qy      147 ThrAla-----LeuLeuArgLeuValGluGlyGlyGlyValArgProValArgLeu 163
Db      685 GAAGCATACCAATGTTCTGAGAGCAATTCAGGTCTGGGCGCGCTCGCATCGTATT 744
Qy      164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
Db      745 AGCGCCGTGAGCAGCATCGAATATGCGGTTTATTACAGCACTGCGCATCTTGCTACT 804
Qy      184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
Db      805 TTTCCTTACGGGCTGACCTGGCAGAAAGAAATGTTACAGTTAGCAACTTTCGGGCTC 864
Qy      204 AlaProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerPro 223
Db      865 TCTTCGCCGATTTAACGCCCTTGGAGCTGCGCATGTCGCGGCACTGTTGCTCAGATCCG 924

```

```

; RESULT 6
; US-10-380-132-7
; Publication No. US20030182679A1
; GENERAL INFORMATION:
; APPLICANT: Suncare GmbH & Co. KGaA
; TITLE OF INVENTION: Improved processes for Vitamin E biosynthesis
; FILE REFERENCE: NAE445/2000
; CURRENT APPLICATION NUMBER: US/10/380,132
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1238
; TYPE: DNA

```

Oy	224	GUVALTYGVLASp11leg1nAgserrAPROALa1a1aSer1a	-----	229
Db	925	CAGCTTTATGCCGACATCATTTATGTCGACAGCGTATATGGCGTTAATCAACGTTAC		984
Oy	240	-----	-----	244
Db	985	TATAAGCGTTTGGCGAGCGAGTATGATGCTCGAGACGAGCGATAGACGAGGCTTATTT		1048
Oy	245	GUVALaLeuValSerPheal1a1aLeuValGlyAspAspProAspArg1a1aPro		264
Db	1045	GACAGTTTCCGACAGGTGAGCAGCTGTTCCGAGTATTCACACAGCGTTTTCAGATGAA		110
Oy	265	GlyArg 266		
Db	1105	AGCCGC 1110		
RESULT 7				
US-10-137-310-1				
; Sequence 1, Application US/10137310				
; Publication No. US20030176675A1				
GENERAL INFORMATION:				
; APPLICANT: Valentim, Henry E.				
; APPLICANT: Mileky, Timothy A.				
; TITLE OF INVENTION: TYRA Genes and Uses Thereof				
; FILE REFERENCE: 16515.147				
; CURRENT APPLICATION NUMBER: US/10/137,310				
; CURRENT FILING DATE: 2002-05-03				
; PRIOR APPLICATION NUMBER: US 60/289,527				
; PRIOR FILING DATE: 2001-05-09				
; NUMBER OF SEQ ID NOS: 12				
; SOFTWARE: PatentIn version 3.0				
; SEQ ID NO: 1				
; LENGTH: 1122				
; TYPE: DNA				
; ORGANISM: Erwinia herbicola				
US-10-137-310-1				
Alignment Scores:				
Pred. No.: 2,01e-13 Length: 1122				
Score: 225.50 Matches: 70				
Percent Similarity: 48.26% Conservative: 41				
Best Local Similarity: 30.43% Mismatches: 94				
Query Match: 13.79% Indels: 25				
Db: 16 Gaps: 6				
US-10-089-514-6 (1-322) x US-10-137-310-1 (1-1122)				
Oy	8	ValValValGVLG1SerG1yAlaVal1G1yG1yMetPheal1g1yLeuLeuArgG1uAla	27	
Db	304	G1GATGTCGGGTGTAGGGCCAGATGGGCGGCTTTGAAAAAATGCTCGGGCTATCA	363	
Oy	28	G1ySerArgTThrLeuValValAspLeuValProProG1yArgProAspAlaCysLeu	47	
Db	364	GGCTACAGCGTTAAAAAGCTGAT-----	387	
Oy	48	ValG1yAspValTThrAlaProG1yProG1yLeuVal1a1a1aLeuArgAspAlaAspLeu	67	
Db	388	-----AAAGAGAGACTCGGCTCAGGCTGAGACTCTGCTCAGCGATGCCGAATG	435	
Oy	68	ValLeuLeuAlaVal1HisG1uAspVal1a1aLeuValAlaVal1a1aProValTThrArgLeu	87	
Db	436	G1GATCATTTAGCGTCGCCATTCACTGACCGAGCAGATGATTCGCCAACTGCCACCACTG	495	
Oy	88	MetArgProG1yAlaLeuLeuValAspThrLeuSerValArgThr---GlyMetAla1a	106	
Db	496	CCGGAA---GATTTGATTTCTGTCTGCATCTCGGGCTCACTAAAAACCGGCTCTCGACGCA	552	
Oy	107	G1uLeuAla1a1aHisAla1aProG1yValG1nHisValG1yLeuAsnProMetPheal1aPro	126	
Db	553	ATGCTGGCGTCCCATTAACGGGCTGTA-----CTGGGCTGATCCGATGTTGGCCG	606	
Oy	127	Ala1a1aG1yMetTThrG1yArgProVal1a1a1aValValTThrArgAspG1yProG1yVal	146	

D	b607	GACAGCGGAGCCTGGCAAAACAAGT-----GTGCTCTGGTGATGGAACAACAACCG	668
O	y147	ThralA-----LeuleuArgleuValGIuglyGIyGIyAlargProValArgleu	163
D	b661	GAAGCGATACAGTGGTTCTGGAGCAGATTCAAGGTGTGGGGGTGCAGTCGATCGATC	720
O	y164	ThrIAGluGlnHIAAsrArgrhTrhrlaAIathGlAlaleuthrhIsalavalIeu	183
D	b721	AGCGCTGTGACATGACCAGAATCATGATTCATTCAGCCGTGGCTCACTTGTACC	780
O	y184	LeuSerPheGlyLeuAlaleuAlaArgleuGIyValAspValArgAlaIeuaIaIathr	203
D	b781	TTCGCTTAAGTCTGTGCATTTAAGCCAAGAACGTAATCTGGATAGCTGGCGCTC	840
O	y204	AlaProProHisGlnValLeuleuAlaLeuAlaArgValIeuglyGIySerPro	223
D	b841	TGCTGCCCATTTACCGGCTTGAACCTGGCGAGTGTGGCGGTTGTTCGCTCAGATCCG	900
O	y224	GIuValIYrGIyAPrIIeGInArgSerAn	233
D	b901	CACCTTAGCGGATATCATCATGTCTTCA	930
 RESULT 8 US-10-634-548-16			
Sequence 16, Application US/10634548			
Publication No. US20040045051A1			
GENERAL INFORMATION:			
APPLICANT: No. US20040045051A1ris, Susan R			
APPLICANT: Lincoln, Kim			
APPLICANT: Abad, Mark Scott			
APPLICANT: Eilers, Robert			
APPLICANT: Hartsuyker, Karen Kindle			
APPLICANT: Hirshberg, Joseph			
APPLICANT: Karunanandaa, Balasubramini			
APPLICANT: Moshiri, Farhad			
APPLICANT: Stein, Joshua C.			
APPLICANT: Valentin, Henry E.			
APPLICANT: Venkatesh, Tyamagondlu V.			
TITLE OF INVENTION: Tocopherol biosynthesis related genes and used thereof			
FILE REFERENCE: Ren-01-125			
CURRENT APPLICATION NUMBER: US/10/634,548			
CURRENT FILING DATE: 2003-08-05			
PRIOR APPLICATION NUMBER: us 60/400,689			
PRIOR FILING DATE: 2002-08-05			
NUMBER OF SEQ ID NOS: 79			
SOFTWARE: PatentIn version 3.2			
SEQ ID NO 16			
LENGTH: 1122			
TYPE: DNA			
ORGANISM: Erwinia herbicola			
US-10-634-548-16			
 Alignment Scores:			
Pred. No.: 2,01e-13 Length: 1122			
Score: 225.50 Matches: 70			
Percent Similarity: 48.26% Conservative: 41			
Best Local Similarity: 30.43% Mismatches: 94			
Query Match: 13.79% Indels: 25			
DB: 18 Gaps: 6			
 US-10-089-514-6 (1-322) x US-10-634-548-16 (1-1122)			
O	y8	ValValValIGlyISerGIVAlaValaIGlyIymcPheaIaGIyleuEurghIuaIa	27
D	b304	GTGATTTGTCGGTGGTAAAGGCGCAKATGGCCGGCTGTTGAATAAAAATGCTCGGCGTATCA	363
O	y28	GIysEraGrTrIeueValaValaspIeuValProProProGIyArGrProAspaIaCySleu	47
D	b364	GGCTACACGGTTAAAAAGCTGAT-----	387
O	y48	ValGIyaSpValThralaPrroGIyProGIuIeuaIaIaIaIeuarghspAlaaspIeu	67
D	b388	-----AAAGAGACTGGCTCAGGCTGAGCTGTGAGACTTGTCTCAGCGATCCGGATG	435

```
QY 68 ValLeuLeuAlaValHisGluSerValAlaLeuLysAlaValAlaProValThrArgLeu 87
|||.....|||
DB 436 GTGATCATTAAGCTGCCGATTCACCTGCAGCAGACAGTATGCCCACTGCCACCACTG 495
QY 88 MetArgProGlyValAlaLeuLeuAlaAspThrLeuSerValArgThr---GlyMetAlaAla 106
|||.....|||
DB 496 CCGGAA---GATTGATTTCTGGTCGATTCGGCGCTCATGTAACAAACCGCGCTTCGAGGCA 552
QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValAlGlyLeuAsnProMetPheAlaPro 126
|||.....|||
DB 553 ATGCTGGCTGCCCATTAACGGCGCTGTA-----CTGGCTGTCATCGATGTTTGGCCCG 606
QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlThrArgAspGlyProGlyVal 146
607 GACAGCGCAGCGCTGCCAACAACAGGTG-----GTGGTCTGGTGTATGGAGACAACCG 660
QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyValArgProValArgLeu 163
|||.....|||
DB 661 GAAGCGTATCAAGTGTCTCGAGCAGATTCAGGTCTGGGGTCCGCTCGCATCGTATTC 720
QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
|||.....|||
DB 721 ACGCGTGTGAGCATGACCAACATGCAATGCATTCAAGCGCTGCCTACTTTCCTACC 780
QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
781 TTCGCTTATGATCGATTTAGCCGAAGAACGTGATTCAGCTGAGCGGCTC 840
QY 204 AlaProProPheHisGlnValLeuAlaLeuAlaArgValLeuGlyGlySerPro 223
|||.....|||
DB 841 TCGTCCCATTTACCGGCTTGAACTGGCGATGCTGGCGGCTTTCCTCGTCAGATCCG 900
QY 224 GluValTyrGlyAspIleGlnArgSerAsn 233
|||.....|||
DB 901 CAACTTATGCGGATATCATATGCTCTCA 930
RESULT 9
US-10-433-556-20
; Sequence 20, Application US/10433556
; Publication No. US20040117872A1
; GENERAL INFORMATION:
; APPLICANT: Matringe, Michel
; TITLE OF INVENTION: Novel Targets For Herbicides And Transgenic Plants Resistant To
; FILE REFERENCE: 5500*120
; CURRENT APPLICATION NUMBER: US/10/433,556
; CURRENT FILING DATE: 2003-06-04
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 20
; LENGTH: 1152
; TYPE: DNA
; ORGANISM: Erwinia herbicola
; NAME/KEY: CDS
; LOCATION: (1)..(1119)
US-10-433-556-20
Alignment Scores:
Pred. No.: 2,07e-13 Length: 1152
Score: 225.50 Matches: 85
Percent Similarity: 42.72% Conservative: 44
Best Local Similarity: 28.15% Mismatches: 119
Query Match: 13.79% Indels: 55
DB: 19 Gaps: 9
US-10-089-514-6 (1-322) x US-10-433-556-20 (1-1152)
QY 8 ValValaValGlyGlySerGlyAlaValAlaGlyGlyMetPheAlaGlyLeuLeuArgGluAla 27
|||.....|||
DB 304 GTGATTTGTGGTGTAAAGGCCAGATGGCGCGCTCTTTGAAATAAATGCTCGGCGTATCA 363
```

```
QY 28 GlySerArgThrLeuValAlaLeuLeuValProProProGlyArgProAspAlaCysLeu 47
|||.....|||
DB 364 GGCTACACCGCTTAAACGCTGGAT----- 387
QY 48 ValGlyAspValThrAlaProGlyProGluLeuAlaAlaAlaLeuArgAspAlaAspLeu 67
|||.....|||
DB 388 -----AAGAGAGACAGCGCTTACAGGCTGAGACCTCTGTCAGGATGCGGAATG 435
QY 68 ValLeuLeuAlaValHisGluSerValAlaLeuLysAlaValAlaProValThrArgLeu 87
|||.....|||
DB 436 GTGATCATTAAGCTGCCGATTCACCTGCAGCAGACAGTATGCCCACTGCCACCACTG 495
QY 88 MetArgProGlyValAlaLeuLeuAlaAspThrLeuSerValArgThr---GlyMetAlaAla 106
|||.....|||
DB 496 CCGGAA---GATTGATTTCTGGTCGATTCGGCGCTCATGTAACAAACCGCGCTTCGAGGCA 552
QY 107 GluLeuAlaAlaHisAlaProGlyValGlnHisValAlGlyLeuAsnProMetPheAlaPro 126
|||.....|||
DB 553 ATGCTGGCTGCCCATTAACGGCGCTGTA-----CTGGCTGTCATCGATGTTTGGCCCG 606
QY 127 AlaAlaGlyMetThrGlyArgProValAlaAlaValAlThrArgAspGlyProGlyVal 146
607 GACAGCGCAGCGCTGCCAACAACAGGTG-----GTGGTCTGGTGTATGGAGACAACCG 660
QY 147 ThrAla-----LeuLeuArgLeuValGluGlyGlyValArgProValArgLeu 163
|||.....|||
DB 661 GAAGCGTATCAAGTGTCTCGAGCAGATTCAGGTCTGGGGTCCGCTCGCATCGTATTC 720
QY 164 ThrAlaGluGlnHisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeu 183
|||.....|||
DB 721 ACGCGTGTGAGCATGACCAACATGCAATGCATTCAAGCGCTGCCTACTTTCCTACC 780
QY 184 LeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThr 203
781 TTCGCTTATGATCGATTTAGCCGAAGAACGTGATTCAGCTGAGCGGCTC 840
QY 204 AlaProProPheHisGlnValLeuAlaLeuAlaArgValLeuGlyGlySerPro 223
|||.....|||
DB 841 TCGTCCCATTTACCGGCTTGAACTGGCGATGCTGGCGGCTTTCCTCGTCAGATCCG 900
QY 224 GluValTyrGlyAspIleGlnArgSerAsnProArgAlaAlaSerAlaArgAlaLeu 243
|||.....|||
DB 901 CAACTTATGCGGATATCATATGCTCTCAAGAGATTA-TCTGGCGCTGATAAAGCTTA 959
QY 244 AlaGluAlaLeu-----ArgSerPheAlaAlaLeuValAlGlyAspAspPro----- 258
960 TTACCAAGCGGTTTGGTAAGCATTCGCTGCGAGCAGGGCGACAACAGCGCTTAT 1019
QY 259 -----AspArgAlaAspAlaProGlyArg 266
1020 CCGCAGCTTTAAACGGGTTGAACAGTGTTCGCGATTCACGAAACGTTCTCTGCTCGA 1079
QY 267 AlaAspAlaPro-----GlyHisProGly 274
1080 AAGCGAAGCTTGTTCGATTCGCGCAATGACAGCGCCCATTAATAAAGCAATCCAGTT 1139
QY 275 GlyCys 276
|||.....|||
DB 1140 GGATGC 1145
RESULT 10
US-10-329-670-1
; Sequence 1, Application US/10329670
; Publication No. US20040018503A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,670
; CURRENT FILING DATE: 2002-12-24
; PRIOR APPLICATION NUMBER: US 09/643,990
```

```

? PRIOR FILING DATE: 2000-08-23
? PRIOR APPLICATION NUMBER: US 08/487,429
? PRIOR FILING DATE: 1995-06-07
? PRIOR APPLICATION NUMBER: US 08/426,787
? PRIOR FILING DATE: 1995-04-21
? NUMBER OF SEQ ID NOS: 1
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 1
? LENGTH: 1830121
? TYPE: DNA
? ORGANISM: Haemophilus influenzae
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (4747)..(4747)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (9921)..(9921)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (10150)..(10150)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36543)..(36543)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (29298)..(29298)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36551)..(36551)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (36636)..(36636)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (40808)..(40810)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44416)..(44416)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44905)..(44905)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (44975)..(44975)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45593)..(45593)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (45732)..(45732)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (47036)..(47036)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51334)..(51334)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
```

```

? LOCATION: (51602)..(51602)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51786)..(51786)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (51805)..(51805)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (53369)..(53369)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65309)..(65309)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (65313)..(65313)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (80024)..(80024)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (100091)..(100091)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (102696)..(102696)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (105121)..(105121)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (107248)..(107248)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (117136)..(117136)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119750)..(119750)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (119924)..(119924)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (120038)..(120038)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (121344)..(121344)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122167)..(122167)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (122336)..(122336)
? OTHER INFORMATION: n equals a, t, g or c
? FEATURE:
? NAME/KEY: misc_feature
? LOCATION: (131340)..(131340)
```

```

OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (13910)..(13910)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145171)..(145171)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (145942)..(145942)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (147197)..(147197)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (150841)..(150841)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152500)..(152500)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (152530)..(152530)
OTHER INFORMATION: n equals a, t, g or c

Alignment Scores:
Pred. No.: 9.5e-09 Length: 1830121
Score: 216.00 Matches: 77
Percent Similarity: 45.26% Conservative: 47
Best Local Similarity: 28.10% Mismatches: 118
Query Match: 13.21% Indels: 32
DB: 17 Gaps: 8

US-10-089-514-6 (1-322) x US-10-329-670-1 (1-1830121)
QY 2 SerGlyPheProArgSerValValValGlySerGlyValValGlyMetPheAla 21
DB 1370739 TCTGATATTACCAAAATGTTATGTTGGCGTTATGGAATTAAGCGCTATTATGGC 1370798
QY 22 GlyLeuLeuArgGluAlaGlySerArgThrLeuValValAlaLeuValProProProGly 41
DB 1370799 CGTTATTATTCGTCATCGCTATCCATTCTATTATTAGT----- 1370840
QY 42 ArgProAspAlaCysLeuValGlyAspValThrAlaProGlyProGluLeuAlaAlaAla 61
DB 1370841 CGCGAAGATGGCGCGGTGAAGATATT----- 1370870
QY 62 LeuArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaVal 81
DB 1370871 TTAGCGAATGCTAGTCGATCGTTTCGTCCTATTATCTCACCTTAGAACAATTT 1370930
QY 82 AlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal--- 100

```

```

DB 1370931 GAACGCTTAAACCTTATTAAACGAAACATGCTACTTGTCAGATTAACTCTGTTAAG 1370990
QY 101 ArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaAlaProGlyValGlnHisValGlyLeu 120
DB 1370991 CGTGAACCGCTACGAAATGCTGAAGTTCAATCGTGTCTGT-----TTAGGTTTA 1371044
QY 121 AspProMetPhe---AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAla 139
DB 1371045 CATCCAAATGTTTGTCAGATATTCGAAAGATGCAAAACAAAGTGTTGTCGTTGAT 1371104
QY 140 ThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGluGlyGlyGlyArg 159
DB 1371105 GGAAGTTTCTCGAAGCTTATGATGATGTTTACCTTGACAAATTAATTTGGGGTCAAAA 1371164
QY 160 ProValArgLeuThrAlaGluGluHisAspArgThrThrAlaAlaThrGlnAlaLeuThr 179
DB 1371165 ATTTATCAACCAATGCGACAGAACGATCATATATATGACTTATATACAGCTTGGC 1371224
QY 180 HisAlaValLeuLeuSerPheGlyLeuAlaLeuAlaArgLeuGlyValAspValArgAla 199
DB 1371225 CATTTTTCGACTTTCGGAATGTTTACCTTCCAAACAGCCCATTTATCTCGCTAAT 1371284
QY 200 LeuAlaAlaThrAlaProProProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeu 219
DB 1371285 TTATTGGCACTTCTTCCCTATTATTCGTTAGTACCTTGATAGTGTGCTTATTT 1371344
QY 220 GlyGlySerProGluValTyArgIle-----GlnArgSerAspProAlaAla 237
DB 1371345 GCGCAAGATCGACAGCTTTACGAGATATTATGATGATACAGAAATTTAGCGGTA 1371404
QY 238 -----SerAlaArgArgAlaLeuAlaGluAlaLeuArgSerPheAlaAlaLeuValGly 255
DB 1371405 ATTGAAACCGTTAAACCAACTTACGACGAGGCTTAACTTCTTT----- 1371449
QY 256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
DB 1371450 -----GAAATATATGATGCTCAAGGTTTATTGATGCT 1371482

RESULT 11
US-10-158-865-1
; Sequence 1, Application US/10158865
; Publication No. US20040203093A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra
; FILE REFERENCE: PB186P2C1D1
; CURRENT APPLICATION NUMBER: US/10/158,865
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g

```

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:

FEATURE:
NAME/KEY: misc_feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (10091)..(10091)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122167)..(122167)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (122336)..(122336)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131340)..(131340)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (131360)..(131360)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (139910)..(139910)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (140398)..(140398)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (142750)..(142750)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
LOCATION: (145058)..(145058)
OTHER INFORMATION: n equals a,t,c, or g
FEATURE:
NAME/KEY: misc_feature
```


	/	LOCATION:(145171) ..(145171)	
/	OTHER INFORMATION:	n equals a,t,c, or g	
/	FEATURE:		
/	NAME/KEY:	misc_feature	
/	LOCATION:	(145942) ..(145942)	
/	OTHER INFORMATION:	n equals a,t,c, or g	
/	FEATURE:		
/	NAME/KEY:	misc_feature	
/	LOCATION:	(147197) ..(147197)	
/	OTHER INFORMATION:	n equals a,t,c, or g	
/	FEATURE:		
/	NAME/KEY:	misc_feature	
/	LOCATION:	(150841) ..(150841)	
/	OTHER INFORMATION:	n equals a,t,c, or g	
/	FEATURE:		
/	NAME/KEY:	misc_feature	
/	LOCATION:	(152500) ..(152500)	
/	OTHER INFORMATION:	n equals a,t,c, or g	
/	FEATURE:		
/	NAME/KEY:	misc_feature	
/	LOCATION:	(152530) ..(152530)	
Alignment Scores:		Pred. No.:	9.5e-09
Score:		Length:	1830121
Percent Similarity:		Matches:	77
Best Local Similarity:		Conservative:	47
Query Match:		Mismatches:	118
DB:		Indels:	32
		Gaps:	8
US-10-089-514-6 (1-322) x US-10-158-865-1 (1-1830121)			
Oy	2	SerGIYPheProAArgSerValValAlGlyGlySerGIYalAlaValAlGlyMetPhea	21
Db	1370739	TCTGTATTATCAAAATTGTTATTGTGGCGGTTATGTAATTAGGCGGTTATTTGCC	1370798
Oy	22	GlyLeuLeuAArgGlAlaGlySerArgThrLeuValAlaAspLeuValProProGly	41
Db	1370799	CGTATTTACGTCGCATCTGGCTATCCAAATTTCTATTTAGAT	1370840
Oy	42	ArgProAAspAlaCysLeuValGlYAspValThraAlaProGlyProGluLeuAlaAla	61
Db	1370841	CGCGAAGTGGCGGCTGGAAGTAT	1370870
Oy	62	LeuArAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAla	81
Db	1370871	TTAGGAGATGCTGATGTCGTATCGTTCCCTCATTAATCTCACCTTGAAACAATT	1370930
Oy	82	AlaProValThraArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerVal	100
Db	1370931	GAACGCTTAAACCTTATTATAAGGAAAACAATGCTACTTGACAGATTAAACCTGTTAAG	1370990
Oy	101	ArgThrGlyMetAlaAspLeuAlaHisAlaHisAlaProGlyValIginHisValGlyLeu	120
Db	1370991	CGTGAACCGCTAGCGAAAAATGCTGAAGTTATCTGCTGTT-----TTAGGTTTA	1371044
Oy	121	AsnProMetPhe---AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAl	139
Db	1371045	CATCCAATGTTTGGTGCAGATATTGCCAAGTAGGCAAAACAAGTGTTCGCTTGAT	1371104
Oy	140	ThraArgAspGlyProGlyValThraAlaLeuLeuArgLeuValGluGlyGlyValArg	159
Db	1371105	GGAGCTTTTCTGAACGTTATGAATGGTTACTTGAGCAAATTCAAAATTTGGGGGCAAAA	1371164
Oy	160	ProValArgLeuThrAlaGluGlnHisAspArgThrAlaAlaThrGlnAlaLeuThr	179
Db	1371165	ATTATCAACAACCAATGCCACAGAACAACATATATATGACTTTATATCAAGCCTTGCGC	1371224
Oy	180	HisAlaValLeuLeuSerPheGlyLeuValAlaLeuAlaArgLeuGlyValAspValArgAla	199
Db	1371225	CATTTTTGCATTTTGGCAATGGTTTACACCTTTCCAAACGCCCATTAATCTCGCAAT	1371284
Oy	200	LeuAlaAlaThrAlaProProHisGlnValLeuLeuAlaLeuAlaArgValLeu	219

```

Db      1371285 TTATTGGCACTTTCTTCCCTTATTATTCGGTTAGTACCTTGGAGATGATGCTCGTTTATTT 1371344
Qy      220 GlyGlySerProGluValTYrGlyAspIle-----GlnArgSerAsnProArgAlaAla 237
Db      1371345 GCGGAAGATGAGAGCGCTTACCGCAGATATTATTATGTAGTAAAGTACGAAAAATTTAGCGGTA 1371404
Qy      238 -----SerAlaArgArgAlaLeuValGluValaLeuArgSerPheAlaAlaLeuValGly 255
Db      1371405 ATTTGAACGCTAAACAAACTTACGACGAAAGCCTTACTTCTTT----- 1371449
Qy      256 AspAspProAspArgAlaAspAlaProGlyArgAlaAspAla 269
Db      1371450 -----GAAATTAATGATCTGCAAGGTTTATTGATGCT 1371482

RESULT 12
US-10-981-687-1
; Sequence 1, Application US/10981687
; Publication No. US20050131222A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Fra
; TITLE OF INVENTION: Thereof, and Uses Thereof
; FILE REFERENCE: P186P2CID12
; CURRENT FILING DATE: 2004-11-05
; PRIOR APPLICATION NUMBER: US 10/158,865
; PRIOR FILING DATE: 2002-06-03
; PRIOR APPLICATION NUMBER: US 09/557,884
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 08/476,102
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40808)..(40810)
; OTHER INFORMATION: n equals a,t,c, or g
; FEATURE:

```

```
/ NAME/KEY: misc_feature
/ LOCATION: (44416)..(44416)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44905)..(44905)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (44975)..(44975)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45593)..(45593)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (45732)..(45732)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (47036)..(47036)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51334)..(51334)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51602)..(51602)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51786)..(51786)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (51805)..(51805)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (55369)..(55369)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65309)..(65309)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (65313)..(65313)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (80024)..(80024)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (100691)..(100691)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (102696)..(102696)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (105121)..(105121)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (107248)..(107248)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
```

```
/ LOCATION: (117136)..(117136)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119750)..(119750)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (119924)..(119924)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (120038)..(120038)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a,t,c, or g
/ FEATURE:
```

Alignment Scores:

Db 43 CTGGTGTGCTGTTGGTCTGGAGGTTGATCCGGTGGTTTCGTTTCCCAAAAGGCTTGGCTGAAGC 102
Oy 28 G|y---SerArgThrLeuVal---ValAspLeuValProProbioglyArgProAspAla 45
Db 103 GGGCTGTCTCGGAAATGGTGGGGGGTGCATCTC---GACCCGCAATGCCCAAGCTGGCG 155
Oy 46 CybLeuValGlyAspValThrAlaProGlyProGluLeuValAlaAlaLeuArgAspAla 65
Db 160 GTCGAGTTGGGGCTGGTGCATCTGTCCGAAGACGACCTGGTGGCGGCTTCCGGGGCGCG 219
Oy 66 AspLeuValLeuLeuAlaValAlaGluAspValAlaLeuValAlaAlaProValThr 85
Db 220 GACGGTATCAGATTGGCGGGTCCGATCTGGCCATGGAAATAATGCTGGCGCTTGGCC 279
Oy 86 ArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAla 105
Db 280 GGCATGTGATCTGGGGCAGCGGATCTCGACGAGTGTGGCAGCCCAAG---GGCATGTG 336
Oy 106 AlaGluLeuAlaAlaHisAlaProGlyValGlnHisValGlyLeuValnProMetPheAla 155
Db 337 GTCGCGGGCGGACGGAAGGTTCCGGC-----GGCATGCCGGGGCGTTTCGTG 384
Oy 126 ProAlaAlaGlyMetThrGlyArgProValAlaAlaVal-----138
Db 385 CCGGGCCATCCGATTCCGGCTCCGACAGAGCGGGGTGGAACTTCCAACTCAGAGCTG 444
Oy 139 -----ValThrArgAspGlyProGlyValThrAla 148
Db 445 TTCCTCGTCGATAAAGTATTCTGACCCCGCTTGGACAGACGATCGGGCGGCTGGCC 504
Oy 149 LeuLeu---ArgLeuValGluGlyGlyArgProValArgLeuThrAlaGluGlu 167
Db 505 GTGGTCGATCCGCTTGGCTGAGCTGGGGCCGCGAGCTCGACACATGCAAGTGCAGACT 564
Oy 168 HisAspArgThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGly 187
Db 565 CACGATGAAGTGGTGGCCCGACCACTGCATCTCCCGAC-----CTCGTGGCTTCGCT 618
Oy 188 Leu---AlaLeuAlaArgLeuGlyValAspValArgAlaLeuAlaAlaThrAlaPro 205
Db 619 CTGTGCGATTCACTGCGCAAGCGCATGAATACTTGAGATCTTCGCTTACGGTGGCGGC 678
Oy 206 ProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySerProGluVal 225
Db 679 GGTTCCTCGTAT-----TTCACAAGATCCCGGAAGCGCACCCGGTATG 723
Oy 226 TyrGlyAspAlaGlnArgSerAsnProArgAlaAlaSerAlaArgAlaLeuAlaGlu 245
Db 724 TGGCATGACATCTTCTCCGCAC-----GCGAGGCTGTCTGGCG 765
Oy 246 AlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArgAlaAspAlaProGly 265
Db 766 AACCTCGAATACATTTTGGACGACCTCGACCGCTTGGCGGACCGCGTGCATGCA-----819
Oy 266 ArgAlaAspAlaProGlyHisProGlyGlyCybAspGlyAlaGlyValnLeuAspGlyVal 285
Db 820 -----GGGAGTGGGAC-----CAATTGTGGGGCTC 846
Oy 286 PheGlyGluLeuArg 290
Db 847 TTCACGGCGCGCCGC 861

Search completed: October 6, 2005, 02:37:03
Job time : 3498.63 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Comphen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: October 5, 2005, 09:03:04 ; Search time 3522.58 Seconds
(without alignments)
3479.464 Million cell updates/sec

Title: US-10-089-514-6
Perfect score: 1635
Sequence: 1 MSGPRSVVVGSGAVGWMF.....HCELPRTLRTDDEGKDR 322

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 34239544 seqs, 19032134700 residues
Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame+.p2n.model -DEV=xlh
-Q/cgn2.1/USPTO.epool.h/US10089514/runat_04102005_105744_8043/app_query.fasta.1.1621
-DB=EST -QPMT=fastlap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blonsum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pcr -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10089514_@CGN_1_1_6050_@runat_04102005_105744_8043 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_hic:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gsa1:*
9: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	202.5	12.4	597	9	AG020402 Homo sapi
2	196.5	12.0	677	9	AG020403 Homo sapi
3	162.5	9.9	851	7	CF823705 EST701087
4	162.5	9.9	875	7	CO022117 EST818210
5	154.5	9.4	848	7	CO003446 EST791781
6	154.5	9.4	900	7	CF823364 EST700746
7	152.5	9.3	973	7	CF881643 tr1c028xo
8	146.5	9.0	821	7	CF682058 CCAHL63TR
9	146.5	9.0	830	7	CF702754 CCAHK69TR

10	146	8.9	859	7	CF819790	CF819790 EST697172
C 11	145.5	8.9	1407	9	CL959323	CL959323 OaIFCC002
C 12	143.5	8.8	969	9	AG061086	AG061086 Pan trogl
C 13	141	8.6	1086	9	AY408045	AY408045 Homo sapi
C 14	141	8.6	1684	3	CR595067	CR595067 full-1eng
C 15	140.5	8.6	1145	9	AG071167	AG071167 Pan trogl
16	139	8.5	820	6	CD616495	CD616495 56075281H
17	138.5	8.5	751	7	CF882803	CF882803 tr1c087xe
18	138.5	8.5	805	6	CB909122	CB909122 tr1c087xe
C 19	138.5	8.5	984	5	B0053468	B0053468 AGENCOURT
C 20	138.5	8.5	991	5	BM220915	BM220915 AGENCOURT
21	137.5	8.4	713	7	CF676551	CF676551 CCAC141TR
22	137.5	8.4	747	7	CF721128	CF721128 CCAH122TR
C 23	136.5	8.3	2176	3	CR603382	CR603382 full1-1eng
C 24	134.5	8.2	993	3	CL958449	CL958449 OaIFCC001
25	133.5	8.2	760	6	CB902069	CB902069 tr1c028xo
26	133.5	8.2	760	7	CF871839	CF871839 tr1c028xo
27	132.5	8.1	1098	7	CK210037	CK210037 FGAS02182
C 28	132.5	8.1	1122	5	BM928388	BM928388 AGENCOURT
C 29	131.5	8.0	697	9	CC727568	CC727568 OaIFCC001
C 30	131	8.0	865	5	BX372494	BX372494 BX372494
31	130.5	8.0	1309	3	CR723473	CR723473 Tetradon
32	130	8.0	812	4	B648802	B648802 EST510421
C 33	129.5	7.9	1116	8	BZ659478	BZ659478 PAC62-164
34	129	7.9	1207	5	B0645487	B0645487 AGENCOURT
35	129	7.9	3405	9	CL960595	CL960595 OaIFCC004
36	128.5	7.9	1215	9	AG092472	AG092472 Pan trogl
37	128	7.8	843	7	CF210859	CF210859 CAB20007
38	128	7.8	904	4	BG757997	BG757997 602715041
C 39	128	7.8	1195	9	AG082926	AG082926 Pan trogl
C 40	128	7.8	1758	9	CL965380	CL965380 OaIFCC012
C 41	127.5	7.8	724	2	AM661060	AM661060 833001B03
C 42	127.5	7.8	1056	9	CNS04UNT	AL293782 Tetradon
C 43	127.5	7.8	1062	4	BM544107	BM544107 AGENCOURT
C 44	127.5	7.8	1402	9	CL507797	CL507797 SAIL_788
C 45	126.5	7.7	2328	9	CL982230	CL982230 OaIFSC047

ALIGNMENTS

RESULT 1
AG020402/c
LOCUS AG020402.1 597 bp DNA linear GSS 09-JAN-2003
DEFINITION Homo sapiens genomic DNA, 21q region, clone: B53P19_BB042(Fw).abl,
genomic survey sequence.
ACCESSION AG020402
VERSION AG020402.1 GI:6046346
KEYWORDS GSS.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE
1 Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
Homo sapiens genomic DNA, chromosome 21q
Published Only in Database (1999)
2 (bases 1 to 597)
Hattori,M., Fujiyama,A., Ishii,K., Toyoda,A., Taylor,T.,
Park,H.-S., Yada,T., Watanabe,H. and Sakaki,Y.
Direct Submision
Submitted (16-OCT-1999) Masahira Hattori, RIKEN Genomic Sciences
Center(GSC) c/o Kitasato University, 1-15-1 Kitasato, Sagami-dera,
Kanagawa 228-8555, Japan (E-mail:hattori@hg.ims.u-tokyo.ac.jp,
Tel:81-42-778-9923, Fax:81-42-778-9924)
LOCATION/Qualifiers
1.597
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q"
/clone="B53P19_BB042(Fw).abl"

LOCUS	CF823705
DEFINITION	CF823705 951 bp mRNA linear EST_01-APR-2006 EST701087 Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb Coccidioides posadaasi cDNA clone CIDAN17 5' end, mRNA sequence.
ACCESSION	CF823705
VERSION	CF823705.1 GI:45929762
KEYWORDS	EST..
SOURCE	Coccidioides posadaai Coccidioides posadasi Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides. 1 (bases 1 to 951)
REFERENCE	Gardner,M.J. and Cole,G.T. Analysts of gene expression in Coccidioides posadaei mycelia and spherules via expressed sequence tags unpublished (2003) Other ESTs: EST701086
AUTHORS	Contact: Gardner MJ The Institute for Genomic Research 9712 Medical Center Drive, Rockville, MD 20850, USA Tel.: 301 838 3519 Fax: 301 838 0208 Email: gardner@igr.org
JOURNAL COMMENT	Seq primer: M13 Reverse,
FEATURES	
source	Location/Qualifiers 1..851

```

FEATURES
    source
        location/Qualifiers
            1. 851
                /organism="Coccidioides posadasii"
                /mol_type="mRNA"
                /strain="C735"
                /db_xref="taxon:199306"
                /clone="CIDAN17"
                /dev_stage="saprobic phase (mycelia)"
                /lab_host="E. coli DH10B, T1 phage resistant"
                /clone_lib="Coccidioides posadasii saprobic phase cDNA
                    library, 2 to 4 kb"
                /note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV
                    Coccidioides posadasii saprobic phase cDNA library, size
                    fractionated cDNA 2 to 4 kb"

```

Alignment Scores:	
Pred. No.:	0.000364
Score:	162.50
Percent Similarity:	44.72%
Best Local Similarity:	30.15%
Query Match:	9.94%
DB:	7
Gaps:	8
Length:	85
Matches:	60
Conservative:	29
Mismatches:	85
Indels:	25
Gaps:	8

US-10-089-514-6 (1-322) X CF823705 (1-851)

QY 5 ProArSerValValValGlySerGlyValValGlyGlyMetPheAlaGlyLeuLeu 24
 Db 260 CCAAGATCCATTGGGATCATGGGATGGGATATGGGAAGATATATCTCAACGCTTG 319
 QY 25 ArgGluAlaGlySerArgThrLeuValValaAspLeuValProProGlyArgProAsp 44
 Db 320 AGCCAAAGCGGT-----TGGACATTATATGCTGGCAACCACTCAACATTATGAC 370
 QY 45 AlaCysLeuValGlyAspValThr-----AlaProGlyProGluLeu 58
 Db 371 ACACTGAGGCAGAAATTCATTCCATCAACAAATATCAATATATATCCACAGCATTTG 430
 QY 59 AlaAlaLeuLeuArgAspAlaAspLeuValLeuLeuAlaValAlaGluAspValAlaLeu 78
 Db 431 GTCTCC-----AGAGTTAGCGACATTCATTATCTACAGTTAGGAGCGAAGCCATCAAC 484
 QY 79 LysAlaValAlaProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeu 98
 Db 485 AAAATTGTGGCGGAGATGGCCCATCAACCAAAAGTTGGCGGATGTGGTGTCAAACT 544
 QY 99 SerValArgThrGlyMetAlaAlaGlyLeuAlaAla-----HisAlaProGly--- 114

```

Db      TCCTGCAAA-----GCCCGAGCTGCTGCTTTGGAACCATCTTCTCGGGAC   595
QY      115 ValGlnHisValGlyLeuAsnProMetCHeaAlaProAlaAlaGlyMetThrGlyAlaPro 134
Db      GTGGAGATTATTTGGTTCATTCCTCATGTGACGTAACCTTAAGGACGGCG   655
QY      135 ValAlaAlaValThrArgAspGlyProGlyValThrAlaLeuLeuArgLeuValGlu 154
Db      656 TTGCTTGATTCAGCACAG-----GCCTCGACGAGAGCCTCAGCTTGTGTGA   706
QY      155 -----GlyGlyGlyValArgProValArgLeuThrAlaGluGluHisAspArg 170
Db      707 AGTGTGTTCGCTCTTTCGAATCCAAATCGTTTACTTCACCGGGAGAGGCGATGATCG   766
QY      171 ThrThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAla 189
Db      767 ATCACAGCGACACTCAGCCCGTGCAGCATGCTGCATTTTAAATGTGGAACTGGC   823

```

RESULT 4	
CO022117	
LOCUS	
DEFINITION	CO022117 875 bp mRNA linear EST 10-JUN-2004
	EST818210 Coccidioides posadasii saprofitic phase cDNA library,
	greater than 4kb Coccidioides posadasii cDNA clone CIBCR06_5' end,
	mRNA sequence.
ACCESSION	CO022117
VERSION	CO022117.1 GI:48546838
KEYWORDS	EST.
SOURCE	Coccidioides posadasii
ORGANISM	Coccidioides posadasii
	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
	Ongenates; mitosporic Ongenates; Coccidioides.

REFERENCE	1 (bases 1 to 875)
AUTHORS	Gardner,M.J. and Cole,G.T.
TITLE	Analysis of gene expression in <i>Coccidioides posadasii</i> mycelia and spherules via expressed sequence tags
JOURNAL	Unpublished (2003)
COMMENT	Other ESTs: EST818209

The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel.: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org
Seq primer: M13 Reverse.

FEATURES	Location/Qualifiers
source	1. .875

```

/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="CIBCR06"
/dev_stage="saprobic phase (mycelia)"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_1ib="Coccidioides posadasii saprobic phase cdna
library, greater than 4kb"
/notes="vector: pExpress 1; Site 1: Not T; Site 2: Eco RV;
Coccidioides posadasii saprobic phase cdna library, size
fractionated cdna > 4 kb"

```

ORIGIN	
Alignment Scores:	
Pred. No.:	0.000377
Score:	162.50
Percent Similarity:	44.72%
Best Local Similarity:	30.15%
Query Match:	9.94%
DB:	7
length:	875
Matches:	60
Conservative:	29
Mismatches:	85
Indels:	25
Gaps:	8

US-10-089-514-6 (1-322) x CO022117 (1-875)

Oy 5 PrcarGserValvalGalglySerglAlavalGlglymcPhealacIylLeulu 24
 |||||::: |||||::: |||||::: |||||
Db 295 CCAAGATCCATTGGGATCATTTGGGTGGCGATATGGGAAGATGTATGCTCAGCGCTTG 354

ORGANISM

Coccidioides posadasii
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
Onygenales; mitosporic Onygenales; Coccidioides.

REFERENCE

1 (bases 1 to 900)

AUTHORS

Gardner M.J. and Cole G.T.

TITLE

Analysis of gene expression in Coccidioides posadasii mycelia and
spherules via expressed sequence tags

JOURNAL

Unpublished (2003)

COMMENT

Other ESTs: EST700745

CONTACT

Contact: Gardner MJ
The Institute for Genomic Research
9112 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208

E-MAIL

Email: gardner@igf.org

SEQ PRIMER

Seq primer: M13 Reverse.

FEATURES

source

1..900
/organism="Coccidioides posadasii"
/mol_type="mRNA"
/strain="C735"
/db_xref="taxon:199306"
/clone="CIDAL10"
/dev_stage="saprobiic phase ('mycelial')"
/lab_host="E. coli DH10B, T1 phage resistant"
/clone_lib="Coccidioides posadasii saprobic phase cDNA
library, 2 to 4 kb"
/note="Vector: pExpress 1; Site_1: Not I; Site_2: Eco RV;
Coccidioides posadasii saprobic phase cDNA library, size
fractionated cDNA 2 to 4 kb"

ORIGIN

Alignment Scores:

Pred. No.: 0.00199 Length: 900
Score: 154.50 Matches: 66
Percent Similarity: 41.70% Conservative: 32
Best Local Similarity: 28.09% Mismatches: 94
Query Match: 9.45% Indels: 44
DB: 7 Gaps: 10

US-10-089-514-6 (1-322) x CF823364 (1-900)

QY 63 ArgAspAlaAspLeuValLeuLeuAlaValHisGluAspValAlaLeuValAlaValAla 82
DB 125 AGAGTTAGCGCATTAATCTAAGTGTAGAGCGCAACCATCAACAAATTTGGCG 164
QY 83 ProValThrArgLeuMetArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThr 102
DB 185 GAGTATGGCCCATCAACCAAGTTGGCGCATTTGTTGCAACTTCTGCAAA--- 241
QY 103 GlyMetAlaAlaGluLeuAlaAla-----HisAlaProGly---ValGlnHisVal 118
DB 242 -----GCCCAAGAGCTGGCTTTTGAACACATCTTCTGGGAGCGTGAATAT 255
QY 119 GlyLeuAsnProMetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaVal 138
DB 296 TCGGTTCAATTCCTCATGACGCAACGTAACCTTAAGACACGCGTTGCTTGAAT 355
QY 139 ValThrArgAspGlyProGlyValThrAlaLeuAspLeuValGlu----- 154
DB 356 CAGCACAAGA-----GCCTCTGACGAGAGCCTCAGGTTTGTGAAGTGTGTTGCG 406
QY 155 GlyGlyGlyArgProValArgLeuThrAlaGluGlnHisAspArgThrThrAlaAla 174
DB 407 TCTTTGGAATCCAAATACGTTTAACCTCAGCGGGAGAGGATGATGAATCAGACGGGAC 466
QY 175 ThrGlnAlaLeuThrHisAlaValAlaLeuLeuSerPheGlyLeuAla----- 189
DB 467 ACTCAGGCGCGTGAACGATGCTGATTTTGAATATGATGGAACCTGCGGCAAGCCAAATAC 526
QY 190 -----LeuAlaArgLeu-----GlyValAspValArgAlaLeuAlaAla 202
DB 527 CAATTCCCGTGGAGATATCTCGATATGTGGGTGAATTTGAA----- 568

QY

203 ThrAlaProProHisGlnValLeuLeuAlaLeuLeuAlaArgValLeuGlySer 222

DB

569 -----AAGTCAAATTAATATACATTCAGCATTAAGTGAATAG 610

QY

223 ProGluValTyrGlyAspIleGlnArgSerAspProAlaAlaSerAlaArgAla 242

DB

611 TGGCATGTCTATCTGAGACTGCGATATGAACCTGCGCCAAAGACAGATCCGTCAG 670

QY

243 LeuAlaGluAlaLeuArgSer---PheAlaAlaLeuValGlyAspAspProAlaArgAla 261

DB

671 TATGCAAGATCAAGTGCACCAATATTCAATATTGATGTAGCGCCATCGTATGA-GCT 729

QY

262 AspAlaProGlyArgAlaAspAlaProGlyHisProGlyGlyCys 276

DB

730 -----TMACGGAGAGTCAAGGCTGCCAGGCGGCTGT 762

RESULT 7

CF881643

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

CF881643 973 bp mRNA linear EST 31-OCT-2003
trio28xo14.b1 T. reesei mycelial culture, version 6 October 2003
Hypocrea jeorina cDNA clone trio28xo14, mRNA sequence.
CF881643
CF881643.1 GI:38136325
EST.
Hypocrea jeorina (anamorph: Trichoderma reesei)
Hypocrea jeorina
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
1 (bases 1 to 973)
Diemer S.E., Dunn-Coleman N., Foeman P., Houfek T.D.,
Teunissen P.J.M., van Solingen P., Dankmeyer L., Mitchell T.K.,
Ward M. and Dean R.A.
Characterization of the protein processing and secretion pathways
in a comprehensive set of expressed sequence tags from Trichoderma
reesei
FEMS Microbiol. Lett. 230 (2), 275-282 (2004)

CONTACT

FUNGAL GENOMICS LABORATORY

NORTH CAROLINA STATE UNIVERSITY

CAMPUS BOX 7251, RALEIGH, NC 27695, USA

TEL: 919-513-0020

FAX: 919-513-0024

E-MAIL: ralph.dean@ncsu.edu

SEQ PRIMER: LT-F1 primer.

LOCATION/QUALIFIERS

1..973

/organism="Hypocrea jeorina"

/mol_type="mRNA"

/strain="OM6a"

/db_xref="taxon:51453"

/clone="trio28xo14"

/dev_stage="mycelial"

/clone_lib="T. reesei mycelial culture, Version 6 October 2003"

/note="Vector: pREP3Y, site 1: Not I/Sal I; Mycelial culture grown from 24 hrs to 6 days with varying Carbon and Nitrogen sources and concentrations."

ORIGIN

Alignment Scores:

Pred. No.: 0.0033 Length: 973
Score: 152.50 Matches: 59
Percent Similarity: 42.86% Conservative: 25
Best Local Similarity: 30.10% Mismatches: 85
Query Match: 9.33% Indels: 27
DB: 7 Gaps: 6

US-10-089-514-6 (1-322) x CF881643 (1-973)

QY 12 GlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGlySerArgThr 31
DB 169 GCGATGGGCGACATGGGCGGATGTACGCCGAAGACTGTGCGAAGCTGCGAGATTC 228

ACCESSION CF702754
 VERSION CF702754.1 GI:41556913
 KEYWORDS EST.
 SOURCE Cryptococcus neoformans var. neoformans (Filobasidiella neoformans var. neoformans)
 ORGANISM Cryptococcus neoformans var. neoformans
 BUKARYOTA; FUNGI; Basidiomycota; Hymenomycetes; Heterobasidiomycetes; Tremellomycetidae; Tremellales; Tremellaceae; Filobasidiella
 REFERENCE 1 (bases 1 to 830)
 AUTHORS Loftus, B.
 TITLE End sequencing of clones from a full length enriched, normalized JEC21 cDNA library
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: CCANR69TF
 CONTACT: Brendan Loftus
 TIGR Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-838-3543
 Fax: 301-838-0208
 Email: crypt@tigr.org
 Seq primer: TR.
 FEATURES
 source
 Location/Qualifiers
 1..830
 /organism="Cryptococcus neoformans var. neoformans"
 /mol_type="mRNA"
 /strain="JEC21"
 /db_xref="taxon:40410"
 /clone="CCANR69"
 /note="Vector: pCMVSPORT6; Site 1: NotI EcoRV. The full length, normalized library was prepared from a variety of conditions using RNA provided by Joseph Hellman and Jennifer Lodge"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.00915 Length: 830
 Score: 146.50 Matches: 72
 Percent Similarity: 43.55% Conservative: 36
 Best Local Similarity: 29.03% Mismatches: 110
 Query Match: 8.96% Indels: 31
 DB: Gaps: 11
 US-10-089-514-6 (1-322) x CF702754 (1-830)
 QY 9 ValValGly-----GlySerGlyAlaValGlyMecPheAlaGlyLeuLeuArgGlu 26
 |||||
 DB 94 GTGCTAGGCATCTCCGGATGGGTGATGGACGATGTATCCAAAGCATTTACATCGC 153
 QY 27 AlaGlySerArgThrLeuValValAspLeuValProProGlyArgProAspAla--- 45
 |||||
 DB 154 GCGGCATCGACACTTCTACGTC-----TGCACAAACCCGACTCTTTC 198
 QY 46 ----CylLeuValGlyAspValThrAlaProGlyProGluLeuAla-----Ala 60
 |||||
 DB 199 GAGGCCCTCGAGAAATTCAAAGGACCGCATCATCTCTCCGCAACGCTCACGCC 258
 QY 61 AlaLeuArgAspAlaLeuValLeuLeuAlaValHisGluAspValAlaLeuValAla 80
 |||||
 DB 259 GTCTCCGCTTTCACATTCATTTATTTCTGTC---GAACTGCTGCGTTGCCGGCT 315
 QY 81 Val-----AlaProValThrArgLeuMetArgProGlyValAlaLeuAlaAsp 96
 |||||
 DB 316 GTCTGAGAGATGACGACCTTCGACTCGGGTG-----GCGCGGTTGTAGTCGG 366
 QY 97 ThrLeuSerValArgThrGlyMetAlaAlaGluLeuAlaAlaHisAlaProGlyValGln 116
 |||||
 DB 367 CAGACGAGTGTGAAGCCCGAGAGAGGAGCATTTGAGAGATGTTACCG-----GAG 420
 QY 117 HisValGlyLeu-----AspProMetPheAlaProAlaAlaGlyMetThrGlyArg 133
 |||||

DB 421 GATGTGGGATTACAGATGTACATAGTTTCATGGCCGACGTGTGACGACAGAGGCCAG 480
 QY 134 ProValAlaAlaValAlaThrArgAspGlyPro-----GlyValThrAlaLeuLeuArg 151
 |||||
 DB 481 CCGCTT--ATTATCATATCATCAAGAGGTCCGAAAGAAATGTGAAGTGTGAAGT 537
 QY 152 LeuValGluGlyGlyGlyArgProValArgLeuThrAlaGluGluHisAspArgThr 171
 |||||
 DB 538 GTTTTATAGTCATTCAGAGTGTATACGTTCTTCTAGCTACGAGATGATTAAGG 597
 QY 172 ThrAlaAlaThrGlnAlaLeuThrHisAlaValLeuLeuSerPheGlyLeuAlaLeuAla 191
 |||||
 DB 598 ACTGCCAACACCCAGGCTGTGACGATCGCGCTTCTCAGTATGGGTACAGATGCGAG 657
 QY 192 Arg-LeuGlyValAspValArgAlaLeuAlaAlaThrAlaProProProHisGlnValle 211
 |||||
 DB 658 AATCTTCATCTTACCCATGGAAACTGACGCCCTACGTCTCCGATTC--GAAAGTGTCAA 716
 QY 211 UleuAlaLeuLeuAlaArgValLeuGlyGlySerProGluValTYrGlyAspIleGlnAr 231
 |||||
 DB 717 GGTCAACATTTACCTCCGATATCTACTCTGCGCAATGACACGTCTACGCCGCTTCCGCT 776
 QY 231 GSerAspProArgAlaAlaSer 238
 |||||
 DB 777 GCTGAACCCGTCGCCCAATCC 798
 RESULT 10
 CF819790 859 bp mRNA linear EST 01-APR-2004
 LOCUS CF819790
 DEFINITION EST697172 Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb Coccidioides posadasii cDNA clone CID4071 3' end, mRNA sequence.
 ACCESSION CF819790
 VERSION CF819790.1 GI:45925821
 KEYWORDS EST.
 SOURCE Coccidioides posadasii
 ORGANISM Coccidioides posadasii
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Onygenales; mitosporic Onygenales; Coccidioides.
 REFERENCE 1 (bases 1 to 859)
 AUTHORS Gardner, M.J. and Cole, G.T.
 TITLE Analysis of gene expression in Coccidioides posadasii mycelia and spherules via expressed sequence tags
 JOURNAL Unpublished (2003)
 COMMENT Other ESTs: EST697173
 CONTACT: Gardner MJ
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301 838 3519
 Fax: 301 838 0208
 Email: gardner@tigr.org
 FEATURES
 source
 Location/Qualifiers
 1..859
 /organism="Coccidioides posadasii"
 /mol_type="mRNA"
 /strain="C735"
 /db_xref="taxon:199306"
 /clone="CID4071"
 /dev_stage="saprobic phase (mycelia)"
 /lab_host="E. coli DH10B, T1 phage resistant"
 /clone_lib="Coccidioides posadasii saprobic phase cDNA library, 2 to 4 kb"
 /note="Vector: pExpress 1; Site 1: Not I; Site 2: Eco RV; Coccidioides posadasii saprobic phase cDNA library, size fractionated cDNA 2 to 4 kb"
 ORIGIN
 Alignment Scores:
 Pred. No.: 0.0106 Length: 859
 Score: 146.00 Matches: 59
 Percent Similarity: 43.78% Conservative: 29
 Best Local Similarity: 29.35% Mismatches: 83
 Query Match: 8.93% Indels: 31

DB: 7 Gaps: 7

US-10-089-514-6 (1-322) x CF819790 (1-859)

```

OY      5  ProAqSseValValValGlyGlySerGlyAlaValGlyGlyMetPheAlaGlyLeuLeu 24
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      268 CCAAAATCCATTGGGATCATCTTGGATGGGTGATATGGAGGATGATATGCTAGCGCTTG 327
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      25  ArgGluAlaGlySerArg-----ThreValVal 34
      ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      328 AGCCAAGCGGGTTGGAGACTTAAAGACCCCTTAATACCGTGGCGTATGATTTACTA 387
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      35  AspleValProProProGlyArgProAspAlaCysLeuValGlyAspValThraPro 54
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      388 GCATTAAATGCGCGACCAAACTTACCAATTATGACACACTGAGGACAGAAATTCATATCC 447
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      55  Gly-----ProGluLeuAlaAlaLeuArgAspAlaAspLeuVal 68
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      448 AT-ACAAATATCATATATTTATCCCAACGACATTGGTCTCCAGAGTTAGCGACTTACT 506
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      69  LeuLeuAlaValAlaGlyAspValAlaLeuGlyAlaValAlaProValThraArgLeuMet 88
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      507 ATCTACAGGTGAGGAGCAGAAAGCCATCAACAAATGTGGGCGGATATAGGCCCATCAAC 566
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      89  ArgProGlyAlaLeuLeuAlaAspThrLeuSerValArgThrGlyMetAlaAlaGlyLeu 108
      ::|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      567 AAAGTTGGCGCATGTTGTGGTGTCGTCMAACTTCTCCGCAA-----GCCCCAGAGCTG 617
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      109  AlaAla-----HisAlaProGly---ValGlnHisValGlyLeuAsnProMetPhe 124
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      618 GCTGCTTTTGGAGAACCATCTTCTCTGGGACGGGAGATTATTTCCGATTCATTCCTCACT 677
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      125  AlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThraArgAspGlyPro 144
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      678 GGACCGAAGGTGAACCCCTTAAGGAGACAGCCGTTGCTGATTCAGACACAGA----- 728
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      145  GlyValThraAlaLeuLeuArgLeuValGlu-----GlyGlyGlyGlyArgPro 160
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      729 GCCTCTGACGAGAGCCCTCAGGTTTGTAAAGTGTGTTGCGTCTTTCGAATCCAAATAC 788
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      161  ValArgLeuThrAlaGlyGluHisAspArgThrAlaAlaThraAlaLeuThrHis 180
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      789 GTTATCTCAACCGGGAGAGGATGATCATCAATCAACGCGACATCAAGCCGCTGACGAT 848
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
OY      181  Ala 181
      |||
Db      849 GCT 851

```

RESULT 11
 CL959323/C 1407 bp DNA linear GSS 21-SEP-2004
 LOCUS OSJFCC00244 Oryza sativa Express Library Oryza sativa (indica
 DEFINITION Oryza sativa (indica cultivar-group) genomic, genomic survey sequence.
 CL959323
 ACCESSION CL959323.1 GI:52373397
 VERSION 1
 KEYWORDS GSS.
 SOURCE Oryza sativa (indica cultivar-group)
 ORGANISM Oryza sativa (indica cultivar-group)
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 Ehrhartoideae; Oryzaceae; Oryza.
 Ma, L., Wang, J., Chen, C., Liu, X., Su, N., Li, L., Wang, X., Cao, M.,
 Jiao, Y., Sun, N., Zhang, X., Bao, J., Sun, D., Zhao, H., Yuan, L.,
 Wong, G. K. S., Deng, X. W. and Wang, J.
 An analysis of transcriptional regulation of the rice genome and
 its comparison to Arabidopsis
 Unpublished (2004)
 CONTACT: Chen Chen
 JOURNAL COMMENT Department of Bioinformatic
 Beijing Institute of Genomics
 Chinese Academy of Sciences, Beijing 101300, China
 TITLE
 JOURNAL
 COMMENT
 Tel: 86-10-80481559

Fax: 86-10-80488676
 Email: chenchen@genomics.org.cn
 Rice genomic sequence.
 Class: exon-trapped.

FEATURES
 source location/Qualifiers
 1..1407 /organism="Oryza sativa (indica cultivar-group)"
 /mol_type="genomic DNA"
 /db_xref="taxon:39946"
 /clone_lib="Oryza sativa Express Library"
 /note="Oryza sativa exon trapped genomic sequences"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.0218 Length: 1407
 Score: 145.50 Matches: 97
 Percent Similarity: 36.00% Conservative: 20
 Best Local Similarity: 29.85% Mismatches: 106
 Query Match: 8.90% Indels: 102
 DB: Gaps: 15

US-10-089-514-6 (1-322) x CL959323 (1-1407)

OY	17	GIYGIWmEcPheAaGlyLeuEaRgLIaGlySerAThrLeuValaAspLeu	36
Db	1225	GGGGGGTGAATTGCT-----CGCCTA	1200
OY	37	ValProProProGlyaRgProaPheAaCysLeuValGIaAspValThraAlaProGlyPro	56
Db	1204	CCGGCCGGCCGCCGCCGCCGACGGGGGGGAGGTT-----GCTCGTTGGCCAGGC--	1150
OY	57	GIuLeuAlaAlaLeuEaRgAspAlaAspLeuValLeuEaAlaAlaHisGIaAspVal	76
Db	1153	-----GCCGCTGCATTA-----TTG	1133
OY	77	AlaLeuAlaValaAlaProValThraRgLeuWmEcAaRgProGlyAlaLeuEaAlaAsp	96
Db	1138	GGCAACGGCCGGCTCGCGGCTTCATTTGGGGTGGCGGCGG-----	1090
OY	97	ThrLeuSerValaRg-----ThrGlyMetAla	105
Db	1096	ACTTCGCTCGAGCTGCTGCTGCTGCTGCGCGCGCTCTTTGATGTGTGCTTGGCATTCG	1030
OY	106	AlaGIuLeuAlaAlaHisAlaProGlyVal-GlnHisValGIuLeuAsnProWmEcPheAa	125
Db	1036	GCtGCATGGCCGGCCGGCCGGCTGCTTCGCCGCCACTTGGGGCTGC-----CCCCGTGGCC	980
OY	125	aPro-----AlaAlaGlyMe	130
Db	979	TCTTCGGCGGGCGCGGCTGCGCGTGTGCCTTCGCTGCTGCTGCCGTCGCTGCTGCT	920
OY	130	cThnRgIaRgProValAlaAlaValaValThraRgAspGlyProGlyValThraAla	148
Db	919	CGCTGTTCCTGCTGCTGCTGCTGCGCGCGCGGACGAGACGAACTTGGCGCTGCTGCT	860
OY	149	-----LeuEaRgLeuValGIu-----GIyGIyGIyGIyArgProValAr	162
Db	859	CTCGTCGCTCAATCTCTGCTGCTGCTTGTGACAAACAGCGCGGTGGGTGGTTCGTCG	800
OY	162	gLeuThraGluGluHisAspArgThrThraAlaThnGlnAlaLeuThnHisAlaVala	182
Db	799	C-----GGCAACACCTCCACGTTGGCGGTATCTTAAGAGAGCTC-----CCCT	755
OY	182	lleuLeuSerPheGlyLeuAlaLeuAlaArgLeuGIyAlaAspValaRgAlaLeuAlaAl	202
Db	754	CCCTCTTTCATATGGGGCGGATCCGCTTCGATGGCTCCACCTCCGACCTTCGCGCG	695
OY	202	aThraAlaProProPheHisGlnValLeuEaAlaLeuEaAlaRgValLeuGIyGlyse	222
Db	694	C-----GGCGTGTGGCTCGCTGCGCGGCTCCTTATCTTGG	653
OY	222	rProGluValaIYrGIyAspIleGlnArgSerAsn-----	233

Db 652 CCCATTGTGGAGGAGGCGCGGCGGAGACTGATCGCGCGCGGCGGCTCCG 593
 QY 234 -----ProArxAlaIaSerAlaArgArxAlaLeuAlaI 245
 Db 592 CGGCGCGGCGGAGGCGCGGCGGCTTTCGCGCGGCGGCGGCGGCTCCG 533
 QY 245 uAlaLeuArgSerPheAlaIaLeuValGlyAspAspProAspArgAlaAspAlaProG 265
 Db 532 CGCG-----CGGAGCTCTCGCTCCCTTGTAGGTGGCGAGCCAGATGTGCGCTCCGTACGCA 476
 QY 265 yAlaAlaAspAlaProGlyHisProGlyGlyCyAspGlyAlaGlyAsnLeuAspGlyVa 285
 Db 475 CCTGGCGAGCGCGACCATGCTCGGAGGCGACCGTGATGCGCTGGCGAGCGACGATCT 416
 QY 285 lPheGlyGlyLeu 289
 Db 415 TGACGGTGAATC 403

RESULT 12
 AG061086/c 969 bp DNA linear GSS 03-NOV-2001
 LOCUS Pan troglodytes DNA, clone: PTB-049A11.F, genomic survey sequence.
 AG061086
 VERSION AG061086.1 GI:16612888
 KEYWORDS GSS.
 SOURCE Pan troglodytes (chimpanzee)
 ORGANISM Pan troglodytes

REFERENCE
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 1 Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE BAC end sequences of library PTB
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 969)
 AUTHORS Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T. D., Yada, T.,
 1 Totoki, Y., Watanabe, H. and Sakaki, Y.
 TITLE Direct Submission
 JOURNAL Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
 and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
 1-7-22 Shuhiro-chou, Tsukumi-ku, Yokohama, Kanagawa 230-0045, Japan
 (E-mail: chimbes@gsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/
 Tel: 81-45-503-9111, Fax: 81-45-503-9170)
 Clones are derived from the chimpanzee BAC library PTB This BAC end
 was generated during the K&D process and may have higher chance of
 clone tracking errors.
 COMMENT PRIMERS
 Sequencing: -21M13
 LIBRARY
 Vector : PKS145
 R.site 1 : SacI
 R.site 2 : SacI
 Location/Qualifiers
 1. 969
 /organism="Pan troglodytes"
 /mol_type="genomic DNA"
 /db_xref="taxon:9598"
 /clone="PTB-049A11.F"
 /sex="male"
 /cell_type="lymphoblast"
 /clone_lib="PTB Chimpanzee Male BAC Library"

ORIGIN
 Alignment Scores:
 Pred. No.: 0.0205 Length: 969
 Score: 143.50 Matches: 75
 Percent Similarity: 36.77% Conservative: 7
 Best Local Similarity: 33.63% Mismatches: 103
 Query Match: 8.78% Indels: 39
 DB: 9 Gaps: 9
 US-10-089-514-6 (1-322) x AG061086 (1-969)

QY 84 ValThrArgLeuMetArgProGlyAlaLeuAlaAspThrLeuSerValArgThrGly 103
 Db 968 GTTGTGTTTGTGCGCGCGG-----GTTCCCT 936
 QY 104 MetAlaAlaGlyLeuAlaAlaHisAlaProGlyValGlnHis--ValGlyLeuAspPro 122
 Db 935 CGGCGCGCGCGCGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 876
 QY 123 MetPheAlaProAlaAlaGlyMetThrGlyArgProValAlaAlaValAlaThrArgAsp 142
 Db 875 TCGGCGCTCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816
 QY 143 GlyProGlyValThrAlaLeuLeuArgLeuValGlyGlyGlyValArg-----Pro 160
 Db 815 GCG-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 768
 QY 161 ValArgLeuThrAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 180
 Db 767 GTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 708
 QY 181 AlaValLeuLeuSerPheGlyLeuAlaLeuAlaGlyLeuGlyValAspValArgAlaLeu 200
 Db 707 GCGTGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 650
 QY 201 AlaAlaThrAlaProProProHisGlyValLeuAlaLeuAlaLeuAlaArgValLeuGly 220
 Db 649 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 608
 QY 221 GlySerProGlyValThrGlyAspLeuGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 240
 Db 607 GGG-----GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 575
 QY 241 ArgAlaLeuAlaGlyAlaLeuArgSerPheAlaAlaLeuValGlyAspAspProAspArg 260
 Db 574 GCGGCGCTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 515
 QY 261 AlaAspAlaProGlyArg-----AlaAspAlaProGlyHisProGlyGlyCyAsp 277
 Db 514 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 455
 QY 278 GlyAlaGlyAsnLeuAspGlyValPheGlyGlyLeuArgArgLeuMetGlyProGlyLeu 297
 Db 454 GGNGGGGGGGGCGCTCGCGCGCGT-----GCGCGCGCGCGCG-----GCGCGCGCGCG 407
 QY 298 AlaAlaGly 300
 Db 406 GCGGCGCGG 398

RESULT 13
 AY408045/c 1086 bp DNA linear GSS 15-DEC-2003
 LOCUS Homo sapiens SUV39H1 gene, VIRUAL TRANSCRIPT, partial sequence,
 genomic survey sequence.
 AY408045
 ACCESSION AY408045.1 GI:39764016
 VERSION GSS.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens
 REFERENCE
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 1 (bases 1 to 1086)
 Clark, A.G., Gianowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
 Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,
 Adams, M.D. and Cargill, M.
 Inferring nonneutral evolution from human-chimp-mouse orthologous
 gene trios
 JOURNAL Science 302 (5652), 1960-1963 (2003)
 PUMED 14671302
 REFERENCE
 AUTHORS Clark, A.G., Gianowski, S., Nielson, R., Thomas, P., Kejariwal, A.,
 Todd, M.A., Tanenbaum, D.M., Civallo, D.R., Lu, F., Murphy, B.,
 Ferriera, S., Wang, G., Zheng, X.H., White, T.J., Snihsy, J.J.,

TITLE Adams, M.D. and Cargill, M.
 JOURNAL Direct Submission
 COMMENT Submitted (16-NOV-2003) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
 This sequence was made by sequencing genomic exons and ordering them based on alignment.

FEATURES
 source Location/Qualifiers
 1.1086
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <1..>1086
 /gene="SOV39H1"
 /locus_tag="HCM3093"

ORIGIN

Alignment Scores:
 Pred. No.: 0.0393 Length: 1086
 Score: 141.00 Matches: 81
 Percent Similarity: 44.57% Conservative: 34
 Best Local Similarity: 31.40% Mismatches: 98
 Query Match: 8.62% Indels: 45
 DB: 9 Gaps: 14

US-10-089-514-6 (1-322) x AY408045 (1-1086)

QY 9 ValValaIGlyGlySerGlyAlaValaIGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28
 Db 926 ATAGTATGGCGGATCCACGGTGTACACGCTCTCCAGTAGTCCAGGTCAAGAGTAGT 867
 QY 29 SerArgThrLeuValaIAspLeuValaProProGlyArgProAspAlaCysLeuVal 48
 Db 866 GGCGCCCTGACGGTGTGATCTGGCCCGCG-----CTCTGCTCTCTC 822
 QY 49 GlyAspValThrAlaProGlyProGluLeuAlaAlaLeuArgAspAlaAspLeuVal 68
 Db 821 -----TGAGGTATGATCTCTCCACGTAATCCATGACAA----- 786
 QY 69 LeuLeuAlaValaHisGluAspValaAlaLeuValaValaAlaProValThrArgLeuMet 88
 Db 785 -----GCTGTTCTTGGCAATCTTCTCCAGGTCGGACGCCACGCCACGCCATCATC 732
 QY 89 Arg-----ProGlyAlaLeuLeuAlaAspThrLeuSerValaArgThrGlyMetAlaAla 106
 Db 731 CGTGGCGAAGATCAGACGTCATATCGATACCTTCTGTACACACG---ATTGGGCA 675
 QY 107 GluLeuAlaAlaHisAlaProGlyValaGlnHisValaGlyLeuAsnProMetPheAlaPro 126
 Db 674 GTCATAGCCGACACGGCAGCGGAGTTGCACTGTGATGGGACGCGCGCTCGAAGCCG 615
 QY 127 -----AlaAlaGlyMetThrGly-----ArgProValaAlaAlaValaVal 139
 Db 614 CACTTGCCCTGTGATCTGTAGCAAACTTGTGACGTAGACGCCCCGGGACGACGCTCC 555
 QY 140 ThrArgAspGlyProGlyValaThrAlaLeuLeu-----ArgLeuValaGluGlyGly 156
 Db 554 AGTGGGTGCCACAGACAGTCTGGCACTGACGCCACACGACCTGGTTGAGGAT 495
 QY 157 GlyGlyArgProValaArgLeuThr-----AlaGluGlnHisAspArgThrThr 172
 Db 494 GCC-----CTCACCAACACGCTACTGATGATTAACGAAAGCCCGCG 450
 QY 173 AlaAlaThrGlnAlaLeuThrHisAlaValaLeuLeuSerPheGlyLeuAlaLeuAlaArg 192
 Db 449 AGGGCCGCTCCAG-----GTCCACCTCATTTCTTAACAGTATCGCTCCAGATGGCTGG 396
 QY 193 LeuGlyValaAspValaArgAlaLeuAlaAlaThrAlaProProHisGlnValaLeuLeu 212
 Db 395 CTGGGATTAAGAGCTC---CTGCTCCACGACGAGGAGGCGCGCT-----CTGCTT 348
 QY 213 AlaLeuLeuAlaArgValaLeuGlyGlySer---ProGluValaTyrGlyAspIle-GlnArg 231
 Db 347 GGCCTTCTGACACGATGATTGGCCAAAGCTTGGGTGACAGGTGCGGGGGGTCTTTGACCG 288

QY 231 gSerAnProAlaAlaSerAlaArgAlaLeuAlaGluAlaLeuArg 248
 Db 287 GTGTGTCGCCCGGAGAGAGCTCTTCTTAAGTCTTGTGAGATGCTGTTAGG 236

RESULT 14
 CR595067/c
 LOCUS CR595067.1
 DEFINITION full-length cDNA clone CS0D1012Y112 of B cells (Ramos cell line)
 Cot 25-normalized of Homo sapiens (human).
 ACCESSION CR595067.1 GI:50475874
 VERSION CR595067.1
 KEYWORDS HTC; CNSLT_cDNA.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE
 AUTHORS Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
 JOURNAL Full-length cDNA libraries and normalization
 REMARK Contact : Feng Liang Email : fliang@liferich.com URL :
 http://fulllength.invitrogen.com/ Invitrogen Corporation 1600
 Faraday Avenue
 2 (bases 1 to 1684)
 Genoscope.
 REFERENCE
 AUTHORS Direct Submission
 JOURNAL Submitted (20-JUN-2004) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 COMMENT 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
 end enriched, double-strand cDNA was digested with Not I and cloned
 into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
 was normalized. Library was constructed by Life Technologies, a
 division of Invitrogen.
 FEATURES
 source Location/Qualifiers
 1.1684
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="CS0D1012Y112"
 /tissue_type="B cells (Ramos cell line) Cot 25-normalized"
 /plasmid="pCMVSPORT_6"

ORIGIN

Alignment Scores:
 Pred. No.: 0.0683 Length: 1684
 Score: 141.00 Matches: 81
 Percent Similarity: 44.57% Conservative: 34
 Best Local Similarity: 31.40% Mismatches: 98
 Query Match: 8.62% Indels: 45
 DB: 3 Gaps: 14

US-10-089-514-6 (1-322) x CR595067 (1-1684)

QY 9 ValValaIGlyGlySerGlyAlaValaIGlyGlyMetPheAlaGlyLeuLeuArgGluAlaGly 28
 Db 926 ATAGTATGGCGGATCCACGGTGTACACGCTCTCCAGTAGTCCAGGTCAAGAGTAGT 869
 QY 29 SerArgThrLeuValaIAspLeuValaProProGlyArgProAspAlaCysLeuVal 48
 Db 866 GGCGCCCTGACGGTGTGATCTGGCCCGCG-----CTCTGCTCTCTC 824
 QY 49 GlyAspValThrAlaProGlyProGluLeuAlaAlaAlaLeuArgAspAlaAspLeuVal 68
 Db 823 -----TGAGGTATGATCTCTCCACGTAATCCATGACAA----- 788
 QY 69 LeuLeuAlaValaHisGluAspValaAlaLeuValaValaAlaProValThrArgLeuMet 88
 Db 787 -----GCTGTTCTTGGCAATCTTCTCCAGGTCGGACGCCACGCCACGCCATCATC 734
 QY 89 Arg-----ProGlyAlaLeuLeuAlaAspThrLeuSerValaArgThrGlyMetAlaAla 106
 Db 347 GGCCTTCTGACACGATGATTGGCCAAAGCTTGGGTGACAGGTGCGGGGGGTCTTTGACCG 288

